VISUAL ACUITY IN NORMAL AND DISEASED EYES USING HIGH-PASS FILTERED OPTOTYPES

NILPA SHAH

BSC MCOPTOM

DEPARTMENT OF VISUAL NEUROSCIENCE

INSTITUTE OF OPHTHALMOLOGY

UCL

DOCTOR OF PHILOSOPHY (PhD)

2018

Declaration

I, Nilpa Shah, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors, Professor Roger S. Anderson and Professor Steven C. Dakin, for the inspiration, expertise and guidance that they have provided during this research leading to the development of the Moorfields Acuity Chart. Their informed contributions to the presentation of this work at conferences and in the published manuscripts has been invaluable.

I am forever grateful to Professor Roger Anderson for sharing his time, thoughts and extensive knowledge with me and for the kindness and patience that he has shown throughout. I am privileged to have had the opportunity to learn from, and work alongside, such an exceptionally encouraging, enthusiastic and entertaining mentor! I am eternally thankful to Professor Steven Dakin for kindly welcoming me into his lab and for all his meticulous work in coding the psychophysical tests without which this work would not have been possible.

Several other people have been instrumental to this research. I am indebted to Dr Tony Redmond for his contributions in the initial planning and data collection for the first study, to Dr Julie Matlach for help with data collection in the third study, and Ms Heather Whitaker and Ms Sarah Dobinson for their help with data collection in the clinical studies. I am grateful to Mr Adnan Tufail and Ms Cathy Egan for their expert advice and recruitment support in the investigations involving patients with age-related macular degeneration. I also appreciate the support and advice provided by Dr Pádraig Mulholland on potential future projects and collaborations related to the work presented in this thesis.

I thank Fight for Sight and the Moorfields Eye Charity (formerly the Special Trustees of Moorfields Eye Hospital) for providing the financial support for this PhD and to the Research and Development departments at Moorfields Eye Hospital and UCL for their advice and assistance. I am also sincerely grateful to Peter Allen of PA Vision Ltd for the expert design and printing support in the production and subsequent distribution of the Moorfields Acuity Chart and UCL Business School for their commercial supervision. I truly appreciate the kindness of all the study participants in sacrificing their time to make this work possible.

I would also like to especially thank my colleagues and friends at Moorfields Eye Hospital, the UCL Institute of Ophthalmology, St Thomas' Hospital, Hynes Optometrists and the London Clinic Eye Centre who have been incredibly accommodating and supportive throughout my PhD. I thank Mr Alistair Laidlaw for introducing me to the world of research years before, arming me with the skills and motivation which have been valuable in this work.

Finally, I give my heartfelt appreciation to my parents, to whom this work is dedicated, for their infinite love, endless care and unlimited encouragement through every single step of this journey.

Abstract

The most common clinical test of visual function is visual acuity (VA) measurement. Acuity tests have evolved slowly, incorporating chart design features aimed at improving measurement accuracy and minimising test variability. However, with current gold standard logMAR charts, measurements can still be affected by testing and scoring methods, possibly attributable to variations in relative legibility between conventional letters. This limits the test sensitivity and specificity in detecting a change in clinical status. Furthermore, conventional letter charts have demonstrated an insensitivity to early visual system neural deficits.

High-pass filtered letters have a design in which low spatial frequencies, where conventional letters typically vary, are removed. Constructed with a dark core and light edges, the mean letter luminance matches their grey background. This results in similar detection and recognition thresholds with foveal viewing in normal subjects such that letters appear to vanish when the resolution threshold is reached. Under extra-foveal viewing, these thresholds are seen to separate, indicating the neural sampling limited nature of resolution.

This thesis investigates the functional characteristics and limits to performance of high-pass filtered letters. In laboratory-based studies, high-pass VA thresholds were found to display lower between-letter threshold variability, be more robust to the number of alternative letter choices and more resistant to optical

degradations including defocus and simulated lens ageing compared to conventional letters. When a novel high-pass chart, the Moorfields Acuity Chart, was employed in clinical studies, it displayed VA scores and variability less affected by termination and scoring rules in normal subjects with uncorrected refractive error, whilst better revealing functional loss in age-related macular degeneration (AMD). Thus, it appears that high-pass letters can be incorporated in a clinical test chart offering lower variability and in which recognition thresholds are better correlated with early neural deficits in AMD, in a task already familiar to patients.

Contents

List of tables

Chapter 1

Chapter 3

Chapter 4

Chapter 6

Chapter 7

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of abbreviations

AMD	Age-related macular degeneration
ANOVA	Analysis of variance
anti-VEGF	Anti-vascular endothelial growth factor
AREDS	Age-Related Eye Disease Study
arcmin	Minutes of arc or arc minute
BSI	British Standards Institute
cd/m^2	Candela per square metre
cpd	Cycles per degree
cpl	Cycles per letter
CRT	Cathode ray tube
CS	Contrast sensitivity
C S1/2	Conventional Sloan letter set 1/2
C N1/2	Conventional New letter set 1/2
D	Dioptre
dB	Decibels
ETDRS	Early Treatment of Diabetic Retinopathy Study
HRP	High-pass resolution perimetry
ICO	International Council of Ophthalmology
LCD	Liquid crystal display
LOA	Limits of agreement
LOCS III	Lens Opacities Classification System III
logMAR	Logarithm of the minimum angle of resolution

m	Metre
MAC	Moorfields Acuity Chart
MAC S1/2	Moorfields Acuity Chart Sloan letter set 1/2
MAC N1/2	Moorfields Acuity Chart New letter set 1/2
ms	Millisecond
MOL	Method of limits
nAFC	Number of alternative forced choices
NAS-NRC	National Academy of Sciences-National Research Council
OTF	Optical Transfer Function
PRL	Preferred retinal locus
pThreshold	Threshold performance level
RGC	Retinal ganglion cell
RPE	Retinal pigment epithelium
SD	Standard deviation
s	Second
SF	Spatial frequency
TRV	Test-retest variability
VA	Visual acuity
VO	Vanishing Optotype
γ	Gamma
μm	Micrometres

Equation symbols

c	y-axis intercept value
CD	Cell density
CI	Confidence interval
C_N	Number of correct letters
d	Pupil diameter
L_{max}	Maximum luminance value
L_{min}	Minimum luminance value
MAR	Minimum angle of resolution
m	Gradient of the slope
n	Sample size
r^2	Coeffcient of determination
R	Radius
SD	Standard deviation
VA_{4m}	Visual acuity at 4m
VA_{1m}	Visual acuity at 1m
x	x term
y	y term
λ	Wavelength of light

1. Introduction

CHAPTER 1

1.1 Background

The importance of vision assessment was recognised even over 1000 years ago. First explained in, 'The Book of Fixed Stars', written by the early astronomer Al-Sufi, the Arab Eye Test was used in ancient Persia by the army to evaluate the vision of warriors and by the desert Bedouins to test the eyesight of their children. Two stars, separated by 0.2 degrees or 12 arc minutes (arcmins), Mizar and the fainter Alcor (or Al-Suha) are located in the handle of The Plough, the 7 brightest stars of the constellation Ursa Major (The Great Bear). The ability to discriminate between these two stars was used as an assessment of vision and has been demonstrated to be remarkably equivalent to the modern day '6/6' measurement. This is after taking into consideration factors such as differences in contrast and atmospheric conditions in addition to the actual angular separation (Bohigian, 2008).

'I show him Al-Suha and he shows me the moon'.

(Arabic proverb 900 CE)

Originating in reference to the Arab Eye test, the proverb above describes the person who only sees the obvious, despite being shown the subtle detail.

In the modern era, visual psychophysics has allowed us to better quantify the relationship between our perceptions and the physical properties of visual stimuli. The most common clinical psychophysical test in the assessment of visual function is visual acuity (VA) measurement - an evaluation of the spatial resolution of the visual system (Westheimer, 1965). VA measurements are important for detecting and monitoring refractive deficiencies and other visual system abnormalities (Thorn and Schwartz, 1990), for occupational (Association of Optometrists, 2015) and medico-legal assessments including defining driving standards (Latham et al., 2015, Currie et al., 2000), for classifying and assessing eligibility for visual impairment registration (Department of Health, 2013) and as an inclusion criterion and outcome measure in research (Beck et al., 2007).

Considerable work has been dedicated to developing VA tests, the ultimate goal being to possess a test that is:

- **Accurate** in quantifying the true state of the visual system in order to monitor disease progression and/or treatment efficacy (Lovie-Kitchin, 1988, Arditi and Cagenello, 1993).
- **Precise/reliable** such that in the absence of clinical change to the visual system, repeat measures of VA in an individual give similar results each time so that test noise is small (Becker et al., 2007, Shah et al., 2011b, Rosser et al., 2004).
- **Sensitive** in correctly detecting optical and neural deficiencies resulting in a low false negative rate (Rosser et al., 2003a, Cousens et al., 2004, Geer and Westall, 1996).

- **Specific** in correctly identifying those without visual abnormalities resulting in a low false positive rate (Cousens et al., 2004, Shah et al., 2011b).

It is important to quantify the performance of any clinical VA test against each of these descriptors.

1.2 History of visual acuity measurement

Figure 1.1 summarises the significant historical stages in the development of VA measurement discussed within this introduction. As cited earlier, the ability to resolve the double star is one of the earliest references to vision assessment in 900 CE. Following this, centuries later, the first ever Optometric book, 'The use of eyeglasses', was written by a Spanish friar, Benito Daça de Valdes, in which a VA test based on the ability to resolve a line of twelve mustard seeds was described (Daza de Valdes, 1623: as cited in Runge, 2000). In an effort to address the need for standardised vision tests, Kuechler, a German Ophthalmologist, developed a set of three charts in 1843, in order to avoid memorisation. Each chart comprised of twelve lines of a single word in traditional Gothic script, progressively reducing in size (Kuechler, 1843: as cited in Bennett, 1986, Colenbrander, 2001). Kuechler is credited with introducing the first scaled chart to assess VA. The invention of the Ophthalmoscope in 1851 resulted in a complete lack of test standardisation with different hospitals devising their own using passages of text. In 1854, Eduard von Jaeger, Professor of Ophthalmology at Vienna, published a set of reading samples in a number of languages including German, French and English, using fonts available at that time in the State Printing House in Vienna (Jaeger, 1865: as cited in Runge, 2000). Jaeger's charts proved to be more popular than Kuechler's, possibly since his charts included at least four sizes of smaller type thus providing a critical test of acuity (as cited in Bennett, 1986). Jaeger's notation for assessing near vision is still used today in the United States. Franciscus Donders, a professor of physiology, used some of these larger type samples as a distant target for his work on refraction and accommodation before turning to co-worker and physician at the Netherland Hospital for Eye Patients, Hermann Snellen, to formerly develop a distance measurement tool. In 1861, the term, '*visual acuity*', was created by Donders, which he defined as '*the ratio between a subject's performance and a standard performance*', (Donders, 1864: as cited in Colenbrander, 2001).

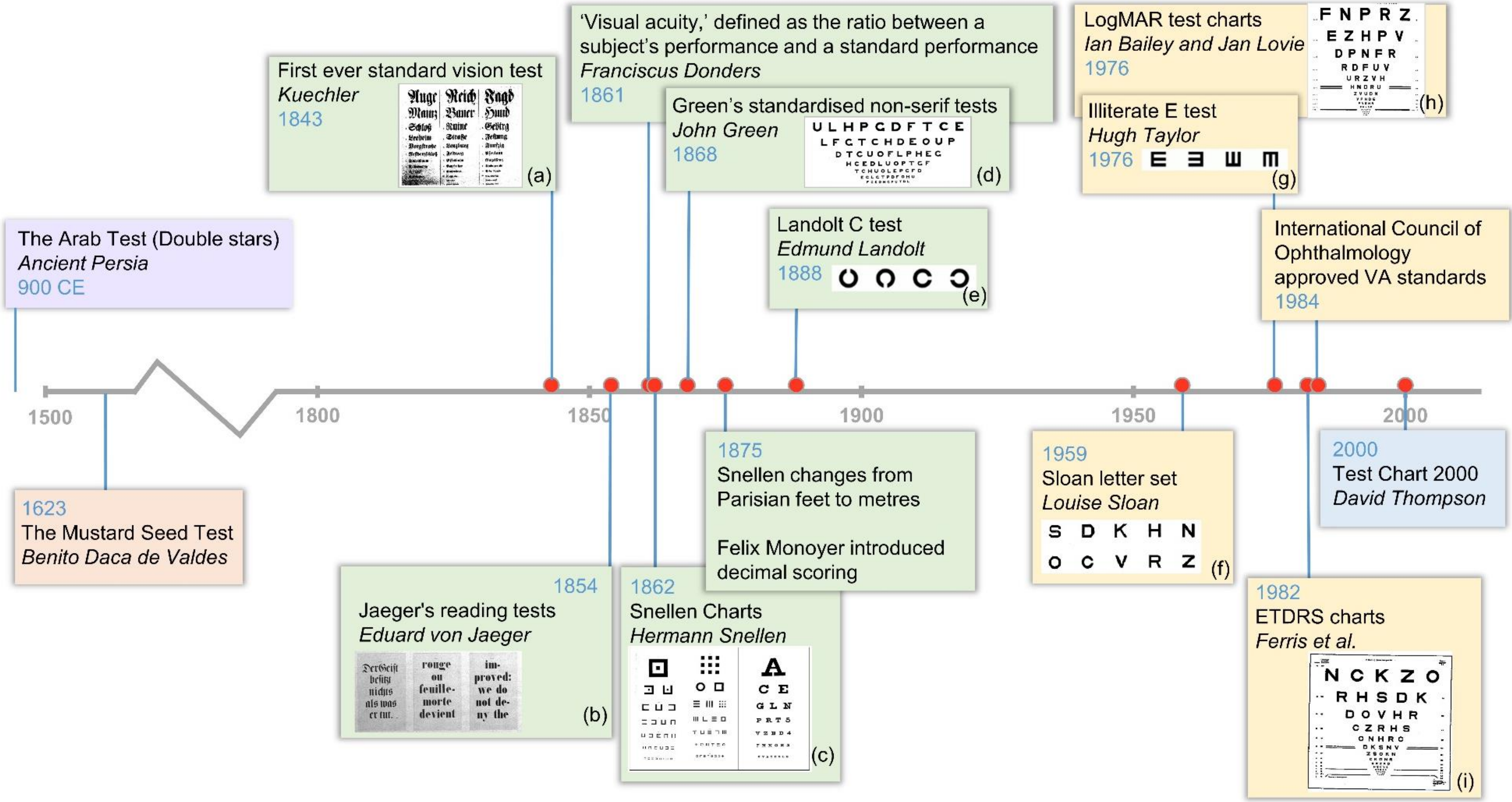


Figure 1.1: Historical landmarks in the development of VA measurement *(adapted from Shah et al., 2016a).*

Test chart images (not to scale) are for illustrative purposes only (*from (a, b) Runge (2000), (c, d, f) Colenbrander (2001), (e, g) Bailey and Lovie-Kitchin (2013), (h) Bailey and Lovie (1976), (i) Ferris et al., (1982).*

1.2.1 The Snellen chart

Hermann Snellen is recognised as being the first to introduce charts with single characters in 1862 (Snellen, 1864). The charts by Snellen were described in, '*Optotypi ad Visum Determinandum*' and became popular internationally very quickly following translation into multiple languages (Colenbrander, 2001). The British Army submitted one of the early large orders when they sought to standardise testing of recruits (Colenbrander, 2001).

Initially experimenting using special symbols, Snellen abandoned these owing to the difficulty that subjects had in accurately describing them. He eventually decided on the use of letters instead which were printed in a serif style Egyptian Paragon font with prominent ornamental cross-strokes at the end of each limb, different in appearance to the optotypes used today (College of Optometrists, 2015). Snellen constructed his optotypes on a 5 unit high x 5 unit wide grid design (although some 5 x 6 letters were used) and defined standard vision as the ability to identify one of his letters when it subtended 5 arcmins, with the detail separated by 1 arcmin. This was based on the discovery, by astronomer Robert Hooke in the 1660s, that the human eye was capable of separating double stars when separated by 1 arcmin. In 1898, Marius Tscherning, a Danish Ophthalmologist reported the inadequacy of using 1 arcmin as a 'normal' value of VA , with many routinely attaining results beyond 6/6 (Tscherning, 1904: as cited in Elliott et al., 1995). Elliott et al., (1995) report average VA measurements of 6/4.8 equivalent in normal subjects free from disease under the age of 60 years old, agreeing with the notion that Snellen's criterion does not represent the normal limits of vision.

Initially calibrated for use at 20 Parisian feet (equivalent to 6.5 metres), Snellen changed from using feet to metres (m) in 1875 (as cited in Bennett, 1986, Jackson and Bailey, 2004) and adopted 6 m as the test distance. In the same year, Ferdinand Monoyer proposed replacing the Snellen fraction with the decimal equivalent. In the present day, the United Kingdom standardises using 6 m, the United States uses 20 feet and Continental Europe uses 5 or 6 m expressed in decimal notation (Colenbrander, 2001). Thus, for example, acuity results are reported as 6/60, 20/200 and 0.1 respectively, all of which are equivalent.

Whilst the development of the Snellen chart represents a significant milestone in the development of standardised VA testing, the limitations and weaknesses of its design are today widely acknowledged and appreciated (Kaiser, 2009, McGraw et al., 1995, Thomson, 2005a, Lovie-Kitchin, 2015). With very few letters at the poor end of the acuity scale (which can be easily memorised), accurate assessment of acuity in low vision patients is severely compromised. Furthermore, the test is not standardised at each VA level. The increasing number of letters on each line down the chart, results in greater contour interaction which has a known detrimental effect on acuity (Levi, 2008) particularly affecting those with amblyopia or central vision problems. This greatly impacts how results can be interpreted such that a two line loss of acuity at one end of the chart is of very different significance to a two line loss at the other end. The lack of systematic spacing between letters and lines and differing steps in letter size also make it difficult to compare results attained at different test distances (Kaiser, 2009) and to detect small changes in acuity with accuracy and sensitivity (Bailey and Lovie, 1976) particularly at the poorer end of the acuity range (McGraw et al., 1995). Despite these limitations, the Snellen chart continues to be widely used and this is attributed to its familiarity, ease of use and short test times (Tewari et al., 2006).

1.2.2 LogMAR charts

Ian Bailey and Jan Lovie of the Kooyong Low Vision Service in Melbourne recognised the importance of designing a VA chart with uniform characteristics throughout, attempting to ensure that under standardised testing conditions, VA is determined merely by the angular subtense of the letter (Bailey and Lovie, 1976, Bailey and Lovie-Kitchin, 2013). In 1976 they introduced new charts, specifically designed for their study investigating vision loss in age-related macular degeneration (AMD) with a test distance of 6 m. Each row, which they determined should have similar legibility, contains an equal number of five letters. For this reason they incorporate the British Standard letter set (D E F H N P R U V Z) (see Section 1.4.1) designed on a 5 unit high x 4 unit wide grid, adopted by the British Standards Institute (BSI) in 1968. The letters are arranged in an attempt to ensure a standard average difficulty between rows using the relative difficulty for the 10 British Standard letters found by Coates and Woodruff as published by Bennett (1965) which is discussed further in Section 1.4.1. These charts have uniform spacing between letters equal to one letter width, and spacing between lines equal to the letter height of the row below. Based on a logarithmic progression in line size using a factor of $\sqrt[10]{10}$ which is equal to 1.2589 or 0.10 log units, the charts have 14 rows allowing for finer acuity grading compared to the Snellen chart. These range from a logMAR (logarithm of the Minimum Angle of Resolution) value of 1.00 logMAR to -0.30 logMAR, indicating that the visual system is resolving detail subtending 10 arcmins (equivalent to a 6/60 letter on the Snellen scale) to 0.5 arcmins (equivalent to 6/3) respectively. This results in uniform contour interaction with a logarithmically equal test step size through the chart. For this reason, testing can be conducted at non-standard distances, for instance when testing low vision subjects, with scores adjusted appropriately using a correction factor. For example, if a test distance of 3 m instead of 6 m is used, a correction factor of 0.30 logMAR should be used to equate the test scores. In direct reference to their structure, these charts are called 'LogMAR charts' (Bailey and Lovie, 1976, Bailey and Lovie-Kitchin, 2013). Figure 1.2 summarises the strengths and weaknesses of Snellen and logMAR chart designs.

Interestingly, in 1868, John Green, Professor of Ophthalmology and Otology at St Louis College of Physicians and Surgeons in Missouri, proposed a chart based on a logarithmic geometric progression (using a factor of $\sqrt[3]{2}$, equivalent to 1.2599) in letter size between lines with proportional spacing between letters (Green, 1868). He also experimented with non-serif Gothic design letters but at that time, faced criticism over their unfinished appearance and for this reason, the chart was not accepted. A century later however, all three of these features were incorporated into the International Standards (Colenbrander, 2001).

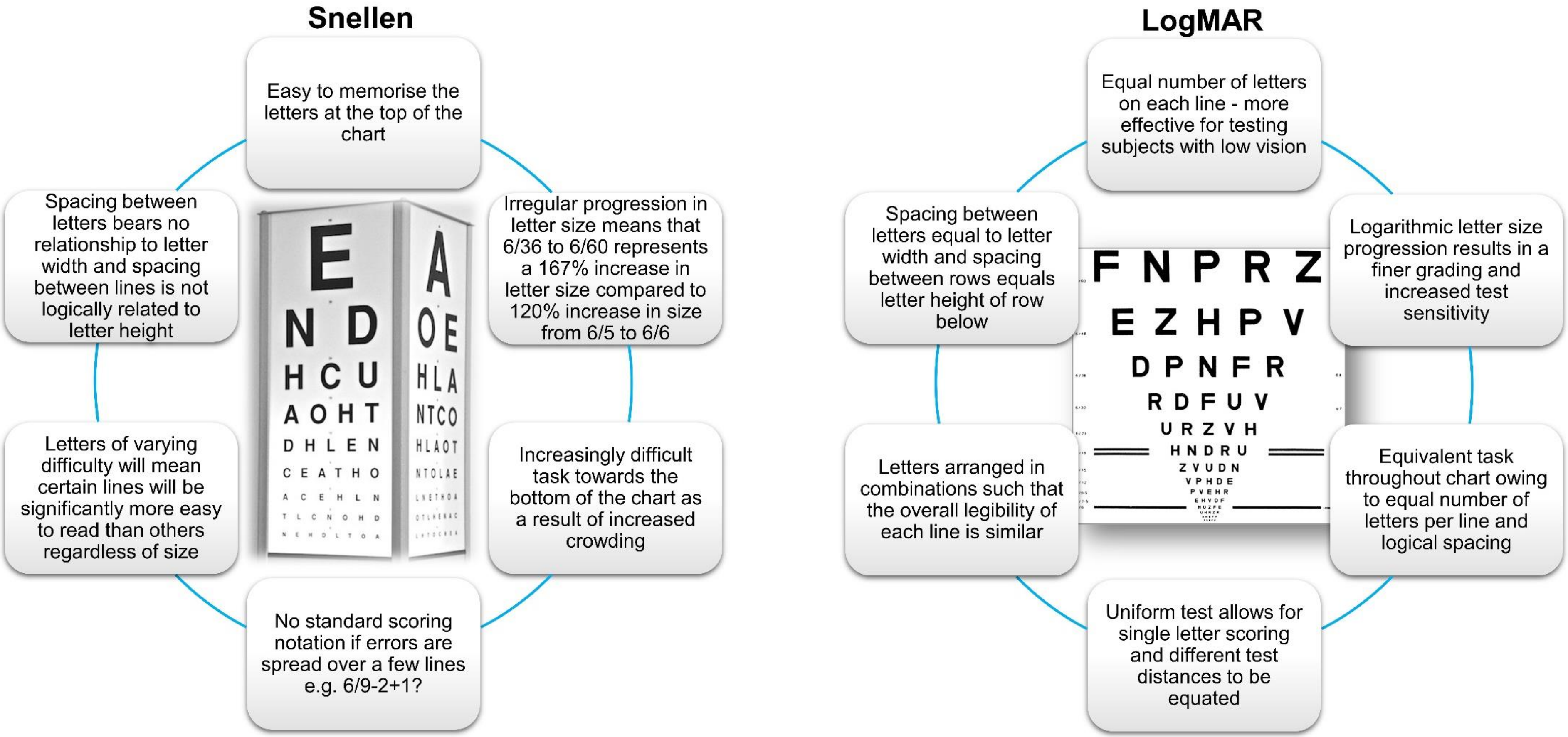


Figure 1.2: Comparison of the strengths and weaknesses of Snellen and logMAR chart designs *(reproduced with permission from Shah et al., 2016a. Visual acuity - more than meets the eye? Optometry Today,(May),69-73).*

Test charts are for illustrative purposes only (not to scale).

1.2.3 The ETDRS chart

In 1982, Rick Ferris et al., of the National Eye Institute established a standardised method of VA measurement for the first phase of the Early Treatment of Diabetic Retinopathy Study (ETDRS) and the Visual Acuity Impairment Study (Ferris et al., 1982). Implementing the chart layout of the Bailey-Lovie logMAR charts, they designed charts which also integrated the recommendations of the Committee on Vision of the National Academy of Sciences-National Research Council (NAS-NRC) Working Group 39 (NAS-NRC Committee on Vision, 1980) by using a 4 m test distance and incorporating the Sloan letter set (C D H K N O R S V Z) using a 5 unit high x 5 unit wide letter grid design. The only deviation from the recommendations is the use of five rather than ten letters per line, with combinations arranged such that each line is of the same average intermediate difficulty (see Section 1.4.2) (Ferris et al., 1982). They named these the 'ETDRS charts' and developed an interpolated scoring protocol in which credit is given to every letter read correctly, termed 'letter-by-letter' or 'single-letter' scoring (see Section 2.2.5 for further details on scoring methods).

1.2.4 'Gold standard' visual acuity testing

In 1984, the International Council of Ophthalmology (ICO) approved a new Visual Acuity Measurement Standard, such that logMAR charts employing single-letter scoring are now considered to be the gold standard tests for VA testing and measurement (International Council of Ophthalmology, 1988). It has however, taken a long time to establish the routine clinical use of logMAR VA testing, being limited only up until recent times to research environments.

Several explanations have been offered as to why this may be (Hussain et al., 2006, Lovie-Kitchin, 2015). These reasons include test familiarity - the Snellen chart has been in use for a long time and the terms '6/6' or '20/20' are well-known and recognised (Hussain et al., 2006, Thomson, 2005b); the scoring system (as described later, see Section 2.2.5) with logMAR charts can also be initially confusing and counterintuitive with negative values reported for better acuity scores (Lim et al., 2010, Thomson, 2005b) and whilst logMAR charts allow for more precise and repeatable VA scores, this comes at the expense of test time (Tewari et al., 2006, Lim et al., 2010). ETDRS chart VA measurements can take twice as long as those taken with the Snellen chart with reported median test times of 60 versus 30 seconds (s) respectively (Laidlaw et al., 2003). Nevertheless, these potential difficulties would be overcome by frequent and repeated use and the lower test variability and increased sensitivity to change of these charts should negate any apprehension to use them.

Falkenstein et al., (2008) and Kaiser (2009) investigated acuity values attained with the ETDRS chart versus the Snellen chart in patients with a range of acuities owing to AMD or a range of other retinal pathology. Both studies found poor agreement between the two with 'better' scores attained on the ETDRS chart particularly at the poorer acuity levels. These studies highlight the caution that should be taken when comparing Snellen acuities obtained in clinic to reported outcomes from clinical trials, which typically employ ETDRS chart testing.

1.2.5 Electronic charts

In line with the incorporation of information technology in all aspects of modern day living, there has been a natural progression to using electronic VA charts in recent years. Test Chart 2000, designed by Professor David Thomson of the Department of Optometry at City University, was the world's first Windows-based computerised test chart (Thomson, 2005b). The numerous advantages of electronic testing include letter randomisation such that the letters can be continuously changed to avoid memorisation (Jackson and Bailey, 2004) and improved control over contrast and illumination can be gained (Ehrmann et al., 2009).

In more recent years, more sophisticated electronic VA measuring systems have become available designed to also control the test procedure. COMPlog (Laidlaw et al., 2008, Shah et al., 2010, Shah et al., 2011b, Shah et al., 2012b, Bokinni et al., 2015) developed by consultant ophthalmologist Mr. Alistair Laidlaw at St Thomas' Hospital in the United Kingdom and the Electronic Visual Acuity Tester (EVA) running the Electronic-ETDRS (E-ETDRS) protocol in the United States (Beck et al., 2003, Cotter et al., 2003) have been extensively validated against the ETDRS chart. They offer automated algorithms allowing for better controlled forced choice testing procedures to measure acuity in a more standardised way. Other benefits of such systems include the ability to measure the whole clinically encountered VA range from one test distance without having to move either the patient or the chart, and automatic test termination criteria and score calculations have helped to overcome some of the fears and resistance to using logMAR acuity testing.

1.3 Visual acuity test limitations

Whilst VA measurements are an essential component of every patient examination, there remain fundamental limitations with the current gold standard tests of acuity which require acknowledgement. Section 1.1 highlighted the desirable qualities of a VA test but numerous inadequacies persist in the accuracy, reliability and in turn sensitivity and specificity which must be considered when making clinical decisions.

1.3.1 Accuracy

VA measurements are affected by the test design and administration as well as the scoring and termination criteria. The more standardised and uniform test design of logMAR charts has greatly improved the accuracy of VA measurements compared to the Snellen chart (Lovie-Kitchin, 1988, Kaiser, 2009, Falkenstein et al., 2008, Lim et al., 2010), particularly in those with low or poor vision. Furthermore, the logarithmic progression allows for different test distances to be used, ensuring that VA scores are not truncated as is often encountered with the Snellen chart. LogMAR charts and, more specifically, ETDRS charts are often used as the reference standard against which the performance of other VA tests are compared. This assumes that the gold standard test is the ideal. However, a study on normal subjects, conducted under strict testing conditions demonstrated that a line of acuity loss on the ETDRS chart, induced by altering the test distance, went undetected in 62% of individuals (Rosser et al., 2003a). Under normal clinical testing conditions, this is likely to be even worse. Furthermore, several studies have shown a reduction of only two or fewer letters in patients with early AMD (Klein et al., 1995) and VA can often remain normal until the fovea is observably affected in even advanced AMD (Wong et al., 2008). This is concerning since high contrast logMAR acuity testing has become the standard for monitoring patients with AMD, with a loss in VA of more than five letters indicating re-treatment with anti-vascular endothelial growth factor (anti-VEGF) intravitreal injections in neovascular AMD in the PrONTO study (Colquitt et al., 2008, Lalwani et al., 2009).

1.3.2 Reliability, sensitivity and specificity

In addition to VA measurement accuracy, test reliability is important. Test-retest variability (TRV) is the noise inherent in a test threshold measurement which accounts for the differences in repeated scores observed in an individual when no true change in clinical status has actually occurred. In VA measurement, the TRV range is often defined by the 95% limits of agreement (LOA) using the methods of Bland-Altman (Bland and Altman, 1986) discussed in Section 2.5.2. These are estimated as the mean difference +/-1.96 x the standard deviation (SD) of the differences between paired measurements if these are approximately normally distributed. A measured VA differing by more than this value on a repeat measure has only a 5% chance of being a false positive or, in other words, not a genuine change in VA. This sets the test specificity at 95% (Rosser et al., 2003a, Cousens et al., 2004). The TRV range is often used to establish the change-criterion whereby a measured value outside this range is considered to represent a genuine and clinically important change. Cousens et al., (2004) demonstrated through their statistical model, supported by empirical data, that a sensitivity of only 50% is achieved in identifying VA changes of similar value to this change-criterion. In order to increase the test sensitivity to 95%, real VA changes of at least 1.84 times the change-criterion are required. It has become apparent over time that by adopting certain strategies, TRV can be reduced (Ricci et al., 1998), which is crucial in improving the test sensitivity to clinically important changes (Rosser et al., 2003a, Cousens et al., 2004).

1.3.3 Reducing test-retest variability

TRV can be influenced by a number of factors. It has been shown to generally increase in the presence of optical defocus (Rosser et al., 2004, Carkeet et al., 2001, Elliott and Sheridan, 1988) and ocular pathology (Patel et al., 2008, Laidlaw et al., 2008). Other influences include the test chart design, test termination criteria, the scoring techniques used, as well as inter-examiner variation (Gibson and Sanderson, 1980). Some of these factors can be minimised by following a standardised testing protocol (Klein et al., 1983) and adopting good testing procedures, e.g. recommendations include measuring VA using the best corrected refractive results. In a group of normal subjects, Rosser et al., (2004) found TRV values ranging from +/-0.11 logMAR for 0 dioptre (D) defocus to +/-0.25 logMAR for 1D of defocus. Several studies have shown the benefit to TRV by taking the mean of multiple repeat measures of VA (Rosser et al., 2003b, Shah et al., 2011b), the disadvantage evidently being the trade off in test time. Testing should be conducted under optimum and consistent test lighting conditions since VA is known to be proportional to chart luminance. Sheedy et al., (1984) recommend that chart luminances should be in the range

80–320 candelas per square metre (cd/m^2) since they found differences of only 0.02 logMAR when the chart luminace is doubled within the range 40–600 cd/m^2. British Standards advise that the minimum background luminance of internally illuminated charts should be a minimum of 120 cd/m^2 (British Standards Institute, 2003).

Forced choice test procedures should be adopted to ensure that subjects are pushed to similar levels each time, reducing differences related to the individual's response criterion or bias and to ensure that termination criteria are fully satisfied. Arditi and Cagnello (1993) suggest that subjects should be required to read the entire chart in order to reap the benefit of the probability of a correct guess. With a two alternative letter forced choice (2AFC) there is a 50% chance of a correct guess compared to 10% with a 10AFC. Carkeet (2001) discussed how this guessing on subthreshold lines can actually introduce further variability to the measurement and describes how, in the worst case scenario, with a 2AFC and termination criteria of five mistakes on a line, a 96.9% chance of proceeding to the subsequent line results in a significant spread of lines where a subject may stop. This transition zone from seeing to non-seeing can be described by probit size (Carkeet et al., 2001) with probit analysis confirming a larger effect on VA of termination criteria and a larger probit size in those with small amounts of optical defocus compared to well corrected subjects. Carkeet (2001) combined exact calculation and Monte Carlo simulation to investigate the effect of the different nAFC available and test termination rules on the mean and SD of logMAR scores, and proposed a number of clinically suitable termination rules for different nAFC. Of importance here are the recommendations made for Bailey-Lovie and ETDRS charts which have a 10AFC (although the observer may not be aware of this) to use a termination rule of four-or-more letters wrong per line for optimum slope corrected SDs.

The influence of scoring techniques on TRV has been examined. With Snellen charts, the line-assignment scoring technique is usually adopted whereby the VA score is taken as the smallest line on which (conventionally), the majority (more than 50%) or 70% (NAS-NRC Committee on Vision, 1980) of the letters are correctly identified. With this technique, patients are given credit for lines and not letters that are correctly identified. Line scoring with the Snellen chart has shown to result in large values of TRV and the Snellen chart is recognised as being a poorly repeatable test (McGraw et al., 1995). With the advent of logMAR design VA charts with rows of equal numbers of letters and systematic changes in line size and spacing, TRV scores significantly improved using the line-assignment scoring technique.

A single-letter scoring protocol was later introduced by Ferris et al., (1982) in the ETDRS study whereby credit is given to each and every single letter read correctly in the final calculated VA score. Each letter is assigned a value of 0.02 logMAR based on the calculation that each line of five letters is equal to 0.10 logMAR. See Section 2.2.5 for details on scoring. Single-letter scoring has resulted in further improved TRV values over the line-assignment technique for logMAR charts by effectively making the grading scale five times finer. The work of Bailey et al., (1991) demonstrates that a coarser grading scale results in an increase in the SD of the discrepancy distribution, in turn extending the

95% confidence limits for change, supporting the notion that a finer grading scale results in an increasing ability of the clinician to detect change in the assessed parameter. Whilst the method of single-letter scoring can be achieved with the Snellen chart, it is much more complex owing to the differing number of letters per line, even though it results in lower TRV values. Table 1.1 illustrates the improved TRV values for logMAR over Snellen charts and further improved TRV using single-letter rather than line-assignment scoring as published by different research groups.

Increasing the number of letters per line whilst employing single-letter scoring can further enhance TRV (Laidlaw et al., 2003, Rosser et al., 2001, Bokinni et al., 2015) with a recent study by Shamir et al., (2016) demonstrating improved reproducibility in VA scores by increasing the number of letters per line up to seven. Raasch et al., (1998) reported that an increase in the number of letters per line by a factor of 'n' can improve precision of VA measurements by a factor of $\sqrt{n}$. This does of course increase test times and a viable balance has to be reached.

Research Group	Subject Group	Snellen Chart		LogMAR Chart	
		Line	Letter	Line	Letter
Rosser et al., (2001)	Cataract, pseudophakia, early glaucoma	+/-0.33	+/-0.24		+/-0.18
Laidlaw et al., (2003)	Amblyopic children	+/-0.30	+/-0.29	+/-0.20	+/-0.14
Lim et al., (2010)	Mixed pathology		+/-0.18		+/-0.14
Elliott and Sheridan (1988)	Normals			+/-0.12	+/-0.07
Bailey et al., (1991)	Normals			+/-0.20	+/-0.10
Vanden Bosch and Wall (1997)	Normals			+/-0.10	+/-0.07
Arditi and Cagenello (1993)	Trained normals			+/-0.13	+/-0.09

Table 1.1: Previously reported TRV values for Snellen and logMAR charts, using either line-assignment or single-letter scoring techniques in different subject groups (all TRV values in logMAR).

Thus in summary, while a lower TRV can be achieved by optimising test conditions and adapting recommended test procedures, it can further be lowered by increasing the number of measurable increments both by:

a) Reducing the step size between lines by using a logMAR rather than Snellen chart. The Snellen chart typically has 9 steps in the range 6/60 to 6/4 where as the logMAR chart has 13 steps for this same VA range.
b) Employing single-letter rather than line-assigment scoring techniques. This results in a grading scale that is five times finer.

Furthermore, the number of alternative letter choices and test termination criteria can also impact on TRV and acuity threshold values and should be considered. It is important to recognise that whilst scores for TRV have improved with the use of logMAR and in particular ETDRS charts, Table 1.1 demonstrates that TRV values of up to 2 logMAR lines can still be observed which has important implications for the monitoring of disease progression and treatment efficacy.

Any useful clinical test should provide a favourable signal-to-noise ratio such that the power of the disease signal provided by the test should not be lost in the background test noise. Whilst adapting VA test design features and adopting standardised testing procedures has improved on this test noise, there appears to be a limit on how much this can be improved. Indeed, Stewart et al., (2006) found no further improvement to TRV on using a randomly interleaved double-staircase technique, using the Sloan letter set in 0.02 logMAR size increments in which the threshold acuity was crossed ten times, to VA measurements attained using a standard ETDRS chart (+/-0.13 versus 0.11 logMAR respectively) in children. The next section looks at the choice of letter styles and sets employed in VA tests and the effects these can have on VA thresholds and variability measures.

1.4 Visual acuity letter sets

Much consideration has been given over the years to the selection and style of letters incorporated into VA test charts. The discussion here does not represent an exhaustive list of all the different approaches to letter choice selection, but aims to highlight some of the key issues and conflicting opinions that have been encountered.

In recognising that not all letters are equally legible, Dennett, an American ophthalmologist investigated the individual recognition distances for each letter of the alphabet and published the results of his experiments in 1885 (Dennett, 1885). This was conducted using normal observers, with corrected refractive error, binocularly viewing a white card with a single non-serif letter constructed on a 5 unit high x 4 unit wide grid with a height of 25 centimetres, placed at the end of a bowling alley. The distance at which the letter could be distinguished from one or two others with which it was most likely to be confused was taken as an indication of its legibility. From this, he produced a test chart in which different letters were scaled in size according to their relative legibilities such that the letters 'A' and 'U' for example were 1.7 x smaller than the letters 'B' and 'S', which were the hardest to distinguish (Dennett, 1885).

In 1919, Hay published details of a test chart based on his notion that the test chart should contain a set of either 'easy' or 'difficult' letters but these should not be mixed (as cited in Bennett, 1965, Banister, 1927) such that if a patient could read one letter on a line, they should be able to read all the letters on that same line, otherwise he assumed they were malingering. Choosing the 'easiest' letters (A C E H L N O T) from his investigations, Hay stressed the importance of ensuring that the design of the letters, all non-serif in style and with a 5 x 5 unit grid construction, was such that they could not be mistaken for another letter. Thus he ensured that the letter 'C' for example had a large break and could not be mistaken for an 'O' (as cited in Bennett, 1965). Whilst Hay selected the letter 'H' as being one of the easier letters to recognise, Hartridge and Owen recommended using letters of medium difficulty in a test chart and identified 'H' as falling within this category (as cited in Banister, 1927). The American Medical Association advised using 18 letters which were classed in four different groups of difficulty and suggested that each line of the chart should use one or two letters from each group, in this case labelling 'H' as being one of the hardest to identify (as cited in Banister, 1927). Observing the incredible variability between studies in reporting the degree of difficulty in identifying certain letters of the alphabet and the arbitrary recommendations of selecting which letters should be incorporated into a VA test, Banister set to explore this further (Banister, 1927). He investigated the recognition of each letter of the alphabet presented tachistoscopically in random order for approximately 0.019 s, noting the incorrect responses which were given for each letter. Figure 1.3 displays the results from Banister's two subject groups showing the order of difficulty in reading each letter compared to the findings of Dennet, Hartridge and Owen and the American Medical Association.

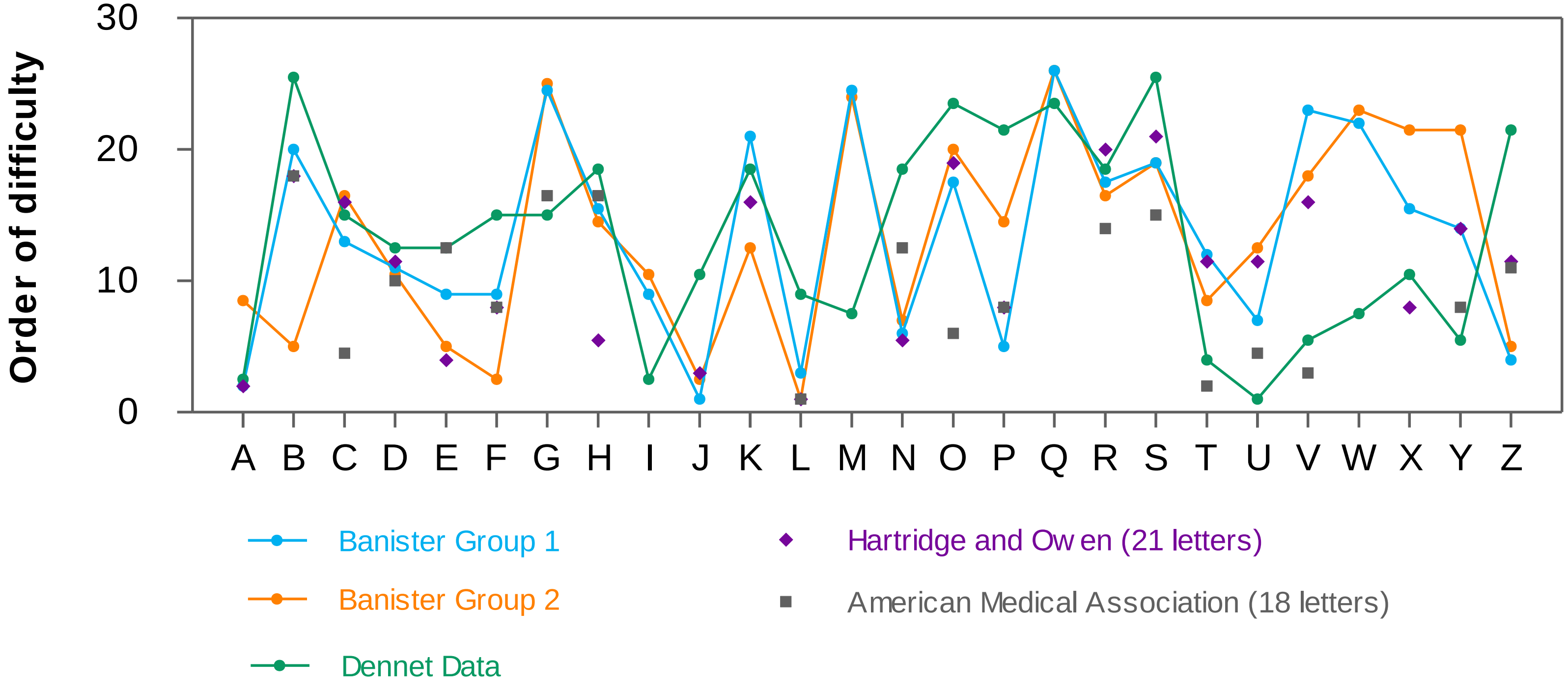


Figure 1.3: Order of letter legibility as reported in various studies *(after Banister, 1927, Dennett, 1885).*

It is evident that while some results agree, there are certain letters which demonstrate large differences between studies which is not completely explicable by the different methodology used. Whilst the letter font type may provide some explanation, Banister resolved that the more likely reason was that letters were incorrectly identified in groups such that when subjects recognised a letter as being from a certain group, they develop some sort of psychological preference to name a certain letter within it. For example, 'C D G O Q' have a very similar outline and form one group as do 'H K M N W' which forms another (Banister, 1927). Banister conducted a second experiment using the ten letters from the 'classes' above and ten 'individual' letters and found that a larger number of letters were read incorrectly from the 'class' group than the 'individual' group. He thus concluded that a VA test should include letters from the same class rather than individual letters as this would more clearly define which letters can be read correctly and those which cannot.

In 1942, Walker discussed his concerns about the Snellen chart (Walker, 1942). He found it irrational that VA's of beyond 6/6 could be attained from a distance of 6 m and felt this demonstrated a fault with the Snellen chart design. Proposing that the problem lay with the relatively little amount of black compared to white on the chart rendering the letters far too easy to see, he introduced a new test type (Figure 1.4). This test used six letters (E F L N T Z) that could be made exactly with 12 to 13 black squares and the corresponding number of white squares in a 5 x 5 unit grid design, presented on a grey background (Walker, 1942).

Figure 1.4: Walker's test type *(reproduced from Walker, 1942)*. Grey scale not correct but for illustrative purposes only.

1.4.1 British Standard letter set

In 1968, the BSI recommended that VA charts use ten letters (D E F H N P R U V Z) selected for their similar legibility based on the work of Coates and Woodruff, known as the British Standard letter set (as cited in Bennett, 1965). These letters, based on a 5 unit high x 4 unit wide grid, non-serif design, with the width of each limb equal to one unit were found to have relative legibilities ranging from 0.91 to 1.09 and were incorporated into the Bailey Lovie logMAR charts (Bailey and Lovie, 1976). Even with their reportedly similar legibilities, the work of McMonnies and Ho (2000) demonstrates that chance combinations of easier or harder letters within this group on a line results in differences in difficulty for lines of the same size between Bailey Lovie charts used for the right and left eyes.

1.4.2 Sloan letter set

The Sloan set was designed initially in 1952 by Louise Sloan, founder of the Low Vision service at the Wilmer Eye Institute of John Hopkins University. These are ten non-serif letters (C D H K N O R S V Z) based on a 5 x 5 unit grid and Sloan proposed to use all ten letters on each line in a VA chart (Colenbrander, 2001). The ten Sloan letters were selected as having similar relative legibility both to each other but also giving similar acuity scores to the Landolt C test (Kniestedt and Stamper, 2003) (see Section 1.4.3) when all ten letters were incorporated on each line in investigations by Sloan et al., (1952). Table 1.2 displays the results of these initial investigations into the percentage of correct responses at threshold (defined as the lowest line on which seven or more letters were read correctly) of these ten Sloan letters in 234 eyes with varying degrees of uncorrected refractive error. It is evident that the letters with curved contours (C D S O) are harder to decipher than the letters composed of straight lines (H K N R V Z).

Letter	Percentage of corrrect responses at threshold
Z	94.0
N	91.6
H	89.3
R	86.3
V	84.6
K	82.1
D	79.5
C	71.4
O	71.0
S	70.6

Table 1.2: **Percentage of correct responses at threshold of the ten Sloan letters in 234 eyes with varying types and degrees of uncorrected ametropia** *(after Sloan et al., 1952).*

In 1980 the NAS-NRC Committee on Vision adopted the Sloan letter set (NAS-NRC Committee on Vision, 1980) which was later incorporated by Ferris et al., (1982) in the ETDRS charts, using only five rather than all ten Sloan letters on each line. Letter combinations were carefully selected using the letter difficulty scores (Table 1.2) quoted by Sloan et al., (1952), such that the average of the accumulated letter difficulty scores for each of the 5 letters on each row is a constant value, similar to the average difficulty score for all 10 letters (82.04%). The maximum difference in average difficulty scores between different lines is less than 1% (Ferris et al., 1982).

Ferris et al., (1993) subsequently conducted their own investigations into the relative legibilities of the Sloan letter set using the data collected from the 3710 subjects who participated in the study qualifying visit of the ETDRS, using charts 1 and 2. Their findings were mostly consistent with those of Sloan et al., (1952) such that the curved letters were found to be more difficult than those with straight lines but found the letters 'O' and 'C' to be harder than the 'S'. The percentage of correct responses at threshold of the most easily legible letter, 'Z' was found to be 84.4% whilst that for the hardest letter, 'C' was 39.3%. Using these new relative letter difficulties, the maximum difference in legibility between lines for either chart 1 or 2 is actually 5.86%. Thus Ferris et al., (1993) designed two new revised test charts with new letter combinations in order to further minimize the difference in relative difficulties between lines, managing to reduce this to 1.32%. They (Ferris et al., 1993) concluded that, "*both the current Early Treatment Diabetic Retinopathy Study charts and revised charts have only small differences in line difficulties and are useful for clinical research purposes*."

On closer inspection, it is perhaps more crucial to note that whilst the average difficulty for each line may be similar, the variation in legibility between letters on each line is substantial in comparison. The top part of each graph in Figure 1.5 demonstrates the average difficulty scores for each line on the original ETDRS charts and the new ones. This was calculated by summing the percent correct at threshold values for each of the Sloan letters found by Sloan et al., (1952) in blue and Ferris et al., (1993) in red and dividing this value by 5. A higher number indicates the more 'easy to read' letters that the line contains. The bottom part of each graph demonstrates the maximum difference in letter legibility for each line calculated as the difference between the highest and lowest values for percent correct at threshold values for each line using values found by Sloan et al., (1952) in blue and Ferris et al., (1993) in red. The higher the value, the greater the difference in letter legibility for that line. It can be appreciated that the average difficulty for each line is similar whilst the within-line difference scores are significantly different and remain so even with the new charts, particularly when considering the results found by Ferris et al., (1993). Using the relative legibility values of Sloan et al., (1952), the largest difference in difficulty scores within a line is 23.4% for the original charts and 23% for the new charts. Using the relative legibility values of Ferris et al., (1993) the largest difference is even greater at 45.1% for both the original and new charts. On reflection, this demonstrates that one line can be harder to read than another, not simply because of the change in letter size but also due to the actual letter choices. Furthermore, two lines of equal size on charts 1 and 2 may still differ in relative difficulty.

Alexander et al., (1997) investigated acuity thresholds for each of the Sloan letters in two naïve subjects and one experienced psychophysical observer and demonstrated a range of 0.14 log units in threshold differences between letters, corresponding to almost a line and a half on the logMAR chart. It is important to consider this spread of legibility of the letters in relation to the size of the interval between lines on a logMAR chart and Alexander et al., (1997) consider this to contribute to reported measurements of test variability.

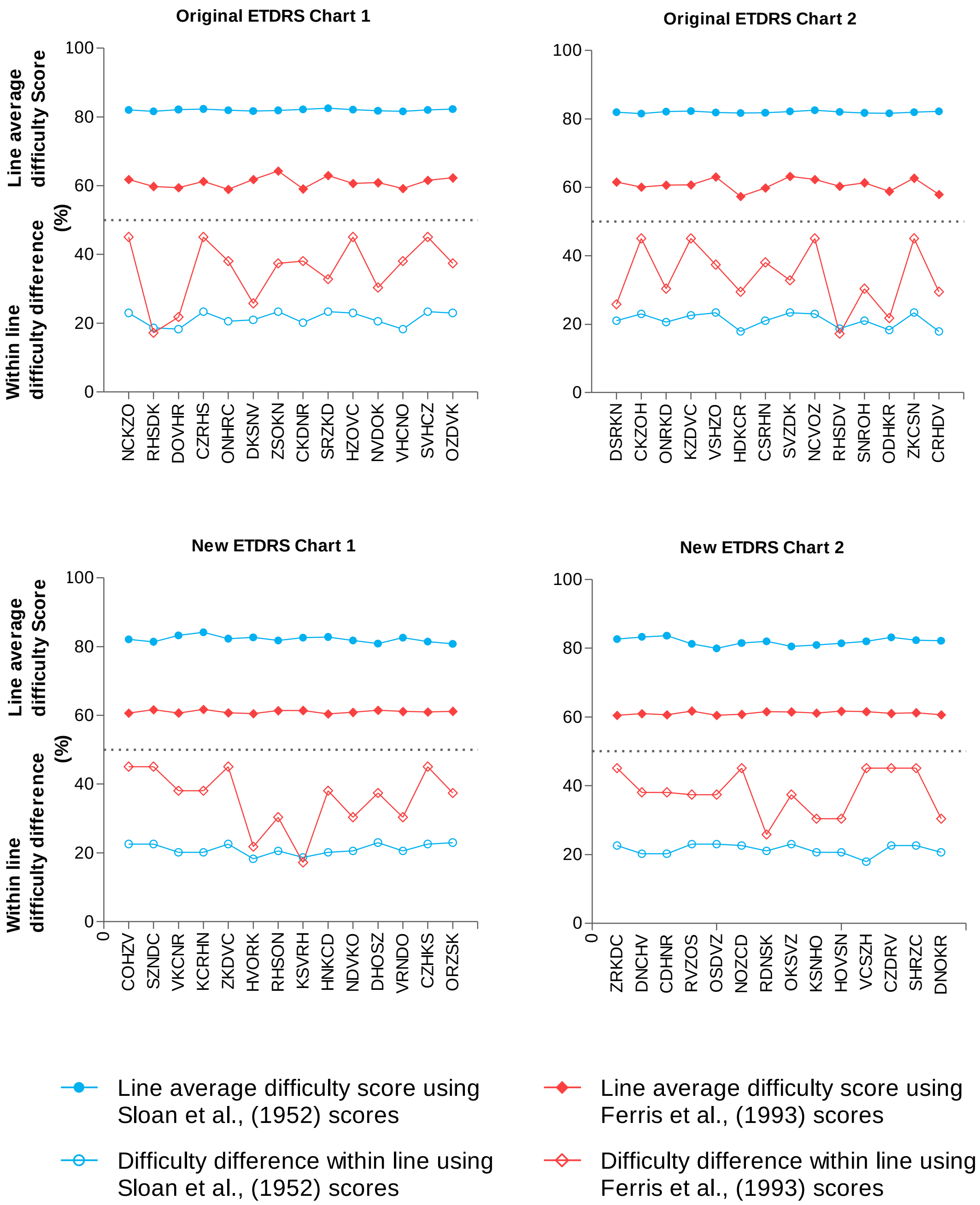


Figure 1.5: Average and difference in difficulty scores per line for the ETDRS charts. The top part of each graph shows the line average difficulty scores and the bottom half, the maximum difference in legibility scores within each line for the original and new ETDRS charts *(after Sloan et al., 1952 and Ferris et al., 1993).* See text for further explanation on calculations.

1.4.3 Landolt C

In acknowledging the difficulty associated with differences in recognisability between different letters, Ophthalmologist Edmund Landolt proposed the Landolt C test in 1888. The Landolt C test is a broken ring, based on a 5 x 5 unit grid, with the stroke width of the letter and the opening measuring one fifth of the diameter. The subject is required to indicate in which position, out of four or eight (if diagonal positions are also used), the opening is. This overcomes the issues associated with differences between letter legibility and the observer's decision making strategy (Jackson and Bailey, 2004). The Landolt C was acknowledged as the gold reference VA target by the ICO.

Several insufficiencies have been identified with this test however, which has resulted in its limited acceptance within the clinical field. The Landolt C test provides only a low number of alternative choices resulting in high guess rates and false-positives, particularly when used in the four positions as favoured by the Visual Functions Committee (International Council of Ophthalmology, 1988). Prince and Fry (1958) also demonstrated that responses can be biased towards a particular orientation in the presence of uncorrected astigmatism and also often to the right, influenced likely by the similarity of the broken ring to the letter C (International Council of Ophthalmology, 1988). Test explanation can be time consuming (Grimm et al., 1994) with ETDRS test times demonstrably shorter (Koenig et al., 2014). Confusion can often arise when presented with multiple optotypes on a line if the patient does not attempt the optotypes in order and subjects have also demonstrated a preference for the ETDRS chart compared to the Landolt C test (Koenig et al., 2014). In addition, there has been a lack of clarity and some inconsistency in reporting the type of acuity, be it resolution or recognition acuity that is measured with the Landolt C test (Wittich et al., 2006, Heinrich and Bach, 2013). Bondarko and Danilova (1995) found that subjects were able to identify the orientation of a Landolt C when the size of the gap corresponded to 66.7 – 55.6 cycles per degree (cpd), significantly beyond the 30 cpd retinal frequency that their subjects could actually resolve using gratings. Further investigation suggests that people can learn to identify an asymmetry in the shape with a flat side indicating the presence of, rather than the actual resolution of the gap (Bondarko and Danilova, 1997).

In recognising the various practical drawbacks of Landolt C acuity assessment, the NAS-NRC committee endorsed the clinical use of other optotypes, provided they would yield corresponding results to those obtained with the Landolt rings (NAS-NRC Committee on Vision, 1980). This was confirmed experimentally for the ten Sloan letters as previously discussed and were adopted by the NAS-NRC Committee. Grimm et al., (1994) looked at the correlation of different optotypes with the Landolt ring in accordance with the International Organisation for Standardisation (ISO) Standard 8597, the relevant standard at the time of their investigation and highlighted the importance of considering individual optotype legibility in addition to the overall mean acuity determined by the optotype set compared to the Landolt ring. Several studies have since demonstrated that mean ETDRS acuity is better than Landolt C acuity by approximately one line (Treacy et al., 2015) and even more so in patients with low vision (Kuo et al., 2011). Landolt C acuity has also been demonstrated to be more affected by optical defocus than letter acuity (Poulere et al., 2013). Simulations by Raasch et al., (1998) found acuity with the British Standard letters to be more similar than the Sloan letters to Landolt C acuity, with better acuities with the letters than Landolt C (mean difference of -0.005 logMAR and -0.038 logMAR respectively), in each case and also more repeatable than either Sloan acuity or Landolt C acuity (SD of 0.036 versus 0.047 logMAR and 0.050 logMAR respectively).

Nevertheless, Landolt C acuity testing remains a useful tool as with the Illiterate E test, designed by Hugh Taylor in 1976, when assessing individuals who are unable to communicate the identity of different letters.

1.5 Letter spatial frequency content

Different models have been proposed to explain the process that the visual system uses for the identification of a stimulus. The template overlap model suggests that recognition of a stimulus image is achieved by comparing it to image templates stored in memory. Hubel and Wiesel's work on the primary visual cortex of the cat and monkey in the 1960's (Hubel and Wiesel, 1962) gave rise to the feature detection model of vision. It was proposed that the visual system analyses the spatial luminance distribution of an image using single neurons in the primary visual cortex which act as line or edge detectors. An alternative model to both of these; the spatial frequency (SF) model was proposed by Campbell and Robson (Campbell and Robson, 1968). Any complex visual image or pattern (such as letters) can also be expressed as a combination of multiple sine wave gratings of different frequencies, orientations, contrast and phase. Fourier analysis allows for an image to be dissected into its constituent sine wave gratings. The SF model suggests that the same neurons that act as line or edge detectors in the feature detection model, actually have a peak sensitivity at a specific range of sine wave frequencies and orientations within an image, analogous to Fourier analysis. Gervais et al., (1984) demonstrated that for letter identification tasks, the SF model is far more convenient to consider than either the template overlap or feature detection models. The Fourier spectra of letter targets contain a broad spectrum of object SFs, measured in cycles per letter (cpl). Indeed, it is for this reason that letters make excellent refraction targets since phase distortions which occur with optical defocus cause artefacts such as contrast reversals and multiple images which, when added to the contrast losses, very quickly cause confusion within the letter making them difficult to recognise (Thorn and Schwartz, 1990, Nestares et al., 2003, Ravikumar et al., 2010). The higher SFs within an image exist as the fine detail such as edges and the low SFs more global information (see Figure 1.6).

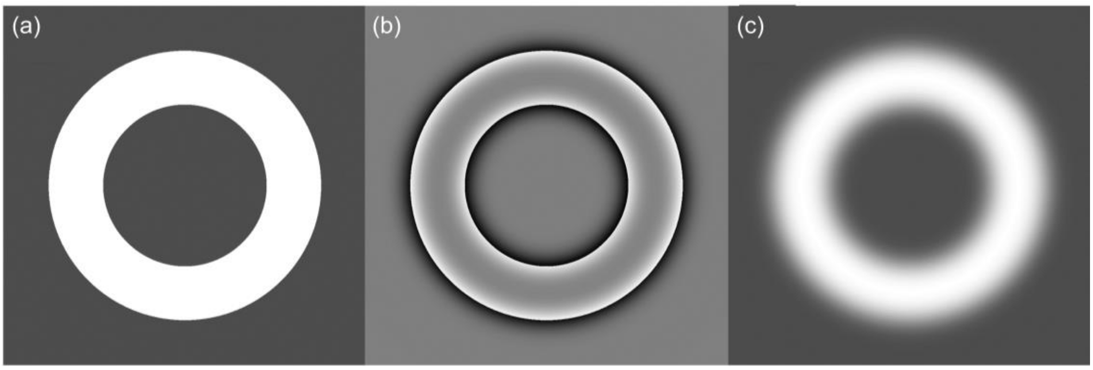


Figure 1.6: Letter SF filtering *(image courtesy of Steven C. Dakin).* The letter 'O' in image (a) has been high-pass filtered (b) to leave only the high SF information which defines the edges of the letter and low-pass filtered (c) to leave the low SF information which gives the overall global information about a letter.

Letter size on gold standard VA charts is specified in terms of the logarithm of the minimum angle of resolution (logMAR) and is usually defined as the finest spatial detail which the visual system can correctly resolve. This scale presumes that the object SFs that determine letter identification are the same across different letter sizes i.e. that scale invariance holds for VA (McAnany et al., 2011). This *object* SF is assumed to be related to the letter stroke width which would be 2.5 cpl or 1 arcmin for the Sloan letters, corresponding to a retinal SF of 30 cpd (McAnany et al., 2011).

The critical object frequency in letter identification has been investigated by several groups and two main approaches have been used in these studies. The first uses a letter filtering approach in which different object frequencies are removed from the letter using spatial filtering. The second is a critical-band noise masking approach which uses spatially filtered luminance noise to mask certain object frequencies. Solomon and Pelli (1994) used filtered noise with different cut-off frequencies to observe a masking effect when presented with letters at around three cpl, independent of letter size. Some have suggested that it is the low SFs that are most important in letter identification (Parish and Sperling, 1991) whilst others have challenged this and proposed that lower SFs do not contribute significantly to this (Howland et al., 1978). Several studies however have confirmed, that only a narrow band of object SFs determine letter identification and that the most important object SFs are actually dependent on letter size (Thorn and Schwartz, 1990, Alexander et al., 1994, Majaj et al., 2002, Chung et al., 2002, McAnany et al., 2011). Alexander et al., (1994) demonstrated a shift in the greatest sensitivity at 2.5 cpl for large and intermediate letter sizes to 0.63 cpl for a small letter size. Majaj et al., (2002) demonstrated this visually very nicely by creating an object (see Figure 1.7) whereby each octave (being a doubling or halving of the SF spectrum) is labelled with a different letter. For example, in Figure 1.7 octave band 3 (2 cpl) was taken from the letter 'C', octave band 4 (4 cpl) from the letter 'D' and so forth to the letter 'F'. It can be seen that as the size of the object changes so does the visible letter, from 'F' at the large size to 'E' at the medium size to 'D' at the smallest size in the bottom right hand corner. This indicates that a different part of the SF spectrum is vital at different sizes. The visual system shifts from using the high object SF content (letter edges) for letter identification when the letters are large to using the low SF content (gross letter strokes) when the letters become too small to resolve the higher object SFs (Majaj et al., 2002, Alexander et al., 1994, Chung et al., 2002). The stimulus energy in these lower SFs continues to decline as the letter is reduced in size until it becomes impossible to discriminate and the VA limit is reached.

Figure 1.7: A demonstration of how the visual system uses different SFs to aid letter identification *(reproduced from Majaj et al., 2002)*. The three objects are identical, except for size, and are created by labelling each octave of SF with a different letter. The visual system shifts from using the high SF content at large letter sizes to progressively lower SF content at smaller object sizes. The largest can be seen as an 'F' (16 cpl), the medium as an 'E' (8 cpl) and the smallest in the bottom right corner as a 'D' (4 cpl).

Alexander et al., (1994) demonstrated that the peak retinal SF remains fairly constant owing to the shift in peak object SF with letter size. They suggest therefore that the reduction in performance observed with decreasing letter size could be attributed to the restricted information that the lower object SFs contain about letter identity (Parish and Sperling, 1991). This shift in the critical object SFs with letter size implies that patients with poorer acuity may be using different critical frequencies compared to those with better acuities. Additionally, Alexander et al., (1997) recognised that relative letter confusions can vary with size as the spectral power of the object SFs differs. Letters particularly vary in their low SF content (Anderson and Thibos, 1999a, Anderson and Thibos, 1999b, Gervais et al., 1984) and this is likely where the differences in legibility arises from at threshold level. Whilst the overall angular subtense of the letters may be the same, some are very similar in their global information whilst others are more different. If different enough, it may be possible to correctly guess the letter identity.

The investigations of Mathew et al., (2011) in assessing the varying difficulty of Snellen letters and common errors revealed that certain letters are frequently confused with others of similar contour. For example, the letters 'B, E and F' were commonly confused as were the circular letters 'O, C, G and D'. Thus, the legibility of a letter will be affected by the number of alternative letter choices and what these letter choices actually are. Reich and Bedell (2000) found that under peripheral viewing conditions, additional letter confusions can occur and a greater range in relative legibility for letters occurs compared to foveal viewing, which has implications when a peripheral retinal locus is adopted, for example in AMD. Raasch et al., (1998) modelled acuity performance using empirical data and demonstrated that most of the variance in an acuity score is attributed to those letters near threshold as certain letters will be identified whilst others will not (Figure 1.8). They investigated the suggestion that choosing more equally legible letters may achieve more reliable results through simulations using the Sloan letter set and a hypothetical letter set in which all the letters were of equal legibility. Surprisingly, their models and simulations revealed a negligible improvement in reliability (SD 0.0214 and 0.0205 logMAR respectively) and they concluded from this that the different legibilities of the letters is not a significant contributor to test variability.

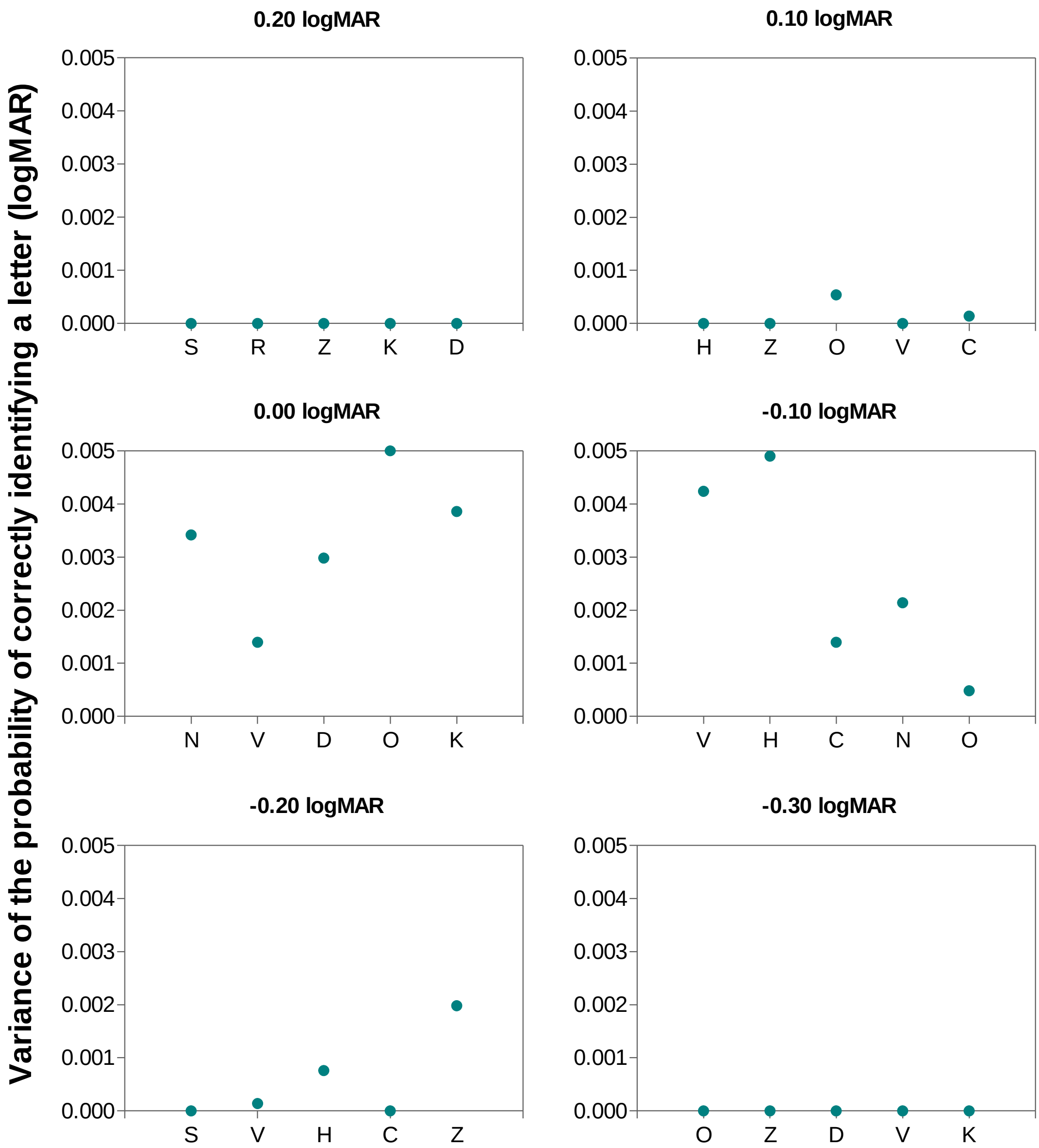


Figure 1.8: Example from simulated acuity measurements of how the variance of the probability of correctly identifying a letter changes. Here the acuity of the observer is set to be approximately 0.00 logMAR and it can be seen that most of the variance in the VA score is contributed by letters near threshold *(after Raasch et al., 1998)*.

1.6 Limits to visual acuity

It is well recognised that different test targets will give rise to different values of VA (Kuo et al., 2011, Elliott and Firth, 2009, Gilbert, 1953, Geer and Westall, 1996, Thorn and Schwartz, 1990). Four main types of VA measurement are recognised (Table 1.3). These are:

Acuity Type	Description
Detection	Determination of the presence of stimulus contrast
Resolution	Differentiation of the separate elements of a visual stimulus
Recognition	Identification of a stimulus
Vernier	Perception of a misalignment of elements

Table 1.3: Summary of different forms of VA measurement.

As seen so far, typical clinically used VA tests consist of letters or optotypes of reducing size with acuity limits represented by some kind of recognition threshold. However, Thibos and Bradley (1993) describe the visual process leading to letter recognition as a three stage hierarchical system of various VA functions in their 'Recognition Pyramid' (Figure 1.9). The ability of the visual system to resolve spatial detail is determined by a combination of optical and neural limiting factors.

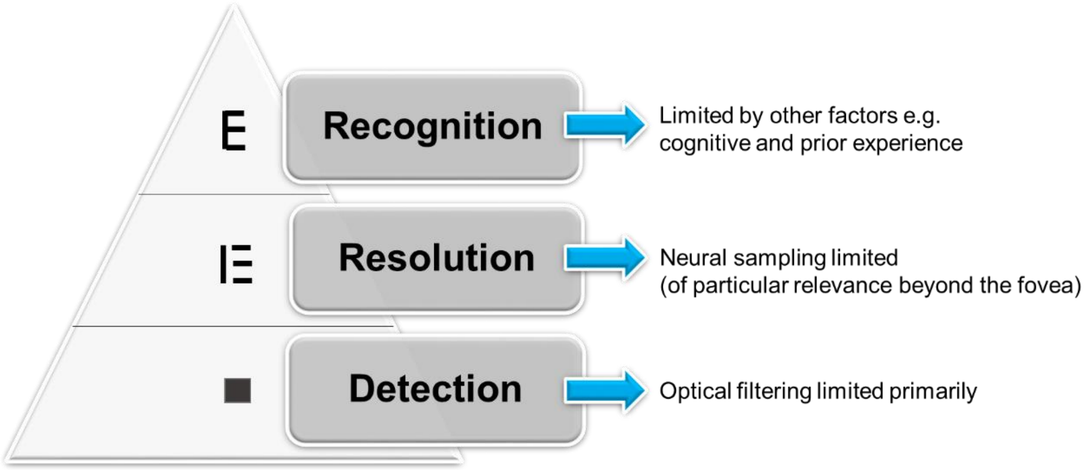


Figure 1.9: Schematic representation of the three stage hierarchical visual processes leading to letter recognition *(adapted from Thibos and Bradley, 1993).*

The initial stage to letter identification is contrast detection and this is limited primarily by the optical filtering effects of the eye. Even in the absence of uncorrected refractive error, a point source is not imaged as such on the retina. Diffraction at the pupil margin results in the imaging of a blurred central circular disc known as 'Airy's disc' surrounded by fainter concentric circular rings on the retina. This distribution of light on the retina is termed the 'point spread function'. The radius (R) of the Airy disc in radians is inversely proportional to the pupil diameter (d) and can be summarised by Equation 1.1 where λ is the wavelength of light.

$$R = 1.22 \lambda / d$$

(Eq. 1.1)

Two point sources will produce two point spread functions, and will only be resolvable if the central maximum in the diffraction pattern of one image coincides with the first minimum of the diffraction pattern of the other image. Put more simply, the two point sources will be resolvable if the Airy discs are separated by at least one radius and this is known as Rayleigh's criterion (Figure 1.10).

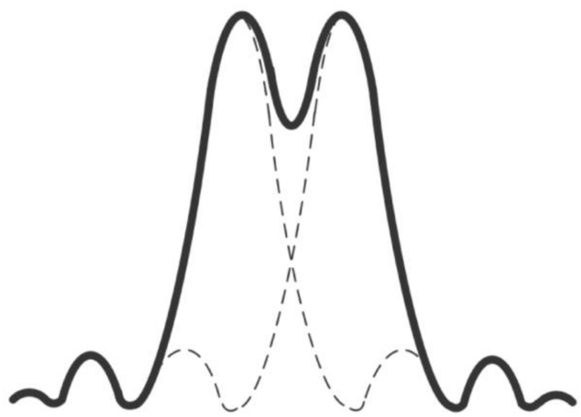


Figure 1.10: Schematic diagram demonstrating Rayleigh's criterion. The dashed lines indicate the illuminance curves for two individual Airy discs and the solid line demonstrates the summation across the area of overlap. In this case, Rayleigh's criteria is satisfied since the peaks of the two are separated by half the diameter of the Airy disc.

Whilst a large pupil may be advantageous at low luminance levels, allowing for maximum retinal illumination and reduced diffraction, increasing pupil size under photopic conditions results in the increase of spherical and other higher order ocular aberrations. The resulting consequence of the eye's optical filtering properties is the attenuation of retinal image contrast for higher SFs to a level below threshold (Schwiegerling, 2000), effectively acting as a low-pass spatial filter.

The second stage of the Recognition Pyramid is resolution which comprises the differentiation of contrast into different letter strokes and is potentially constrained both by the eye's optical quality and neural sampling limit. In order to resolve detail, the image on the retina has to be adequately sampled and processed by visual neurons and this is determined by the size and spacing of photoreceptors for foveal viewing and by variations in photoreceptor to higher-order neuron convergence for peripheral viewing. Helmholtz's theory, originally referring to the discrimination of two points, states that there must be at least one unstimulated neuron between two stimulated neurons for correct discrimination to occur (Helmholtz, 1867: as cited in Green, 1970). The foveolar contains the highest packing density of cones, which have a diameter of approximately 2 to 2.5 micrometres (μm). Thus, assuming the centre to centre distance between cones is between 2 and 3 μm, and the secondary nodal distance for the emmetropic eye is 16.67 millimetres, foveal VA can be expected to reach between 60 and 75 cpd (-0.30 and -0.40 logMAR) (Applegate, 2000). The experiments of Campbell and Green (1965) using interference techniques to bypass the optics of the eye and image gratings directly on the retina, confirmed a cut-off frequency close to 75 cpd. In reality, such resolution levels are not actually achieved under foveal viewing conditions, due to the limiting, optical filtering effects of the eye.

Whilst the eye's optical quality falls with eccentricity from the foveola, the neural spatial resolving ability of the retina declines at a faster rate (Green, 1970).

Westheimer demonstrated an 80% reduction in VA at an eccentricity of 10 degrees compared to foveal fixation (Westheimer, 1987: as cited in Kolb et al., 1995). Green (1970) demonstrated a close correlation between cone packing density and photopic resolution for up to approximately 2 degrees eccentricity, beyond which VA is worse than predicted by cone spacing. Unlike in the fovea in which densely packed cones connect in a 1:1 ratio with retinal ganglion cells (RGCs), resulting in a high sampling frequency, with increasing distance from the fovea, photoreceptors are less densely packed with additional receptor pooling by higher order neurons.

If neural sampling limits resolution, then it is possible for an object to be detected but not veridically resolved. When neural under-sampling occurs, the stimulus can be misrepresented in a perception called aliasing. Figure 1.11, adapted from Thibos (1998), demonstrates how this applies to a grating stimulus, the basic building block of letters. The figure shows two-dimensional neural sampling of a grating with the top row over-sampled, the middle row critically-sampled and the bottom row under-sampled. The middle column displays the perceived orientation of the grating. The strength of the neural response is coded by the luminance of the circles at each sample point with the dashed lines interpolating between the maximum responding neurons. It can be seen here that, whilst the orientation of the sinusoidal grating is represented correctly in the over- and critically-sampled cases, aliasing occurs in the under-sampled scenario to misrepresent the orientation of the grating. The last column displays the perceived SF of the grating. Here the strength of the neural response is represented by the height of the solid lines with the dashed lines interpolating between these responses. Once again, it can be seen that the frequency of the grating is correctly represented in both the over- and critically-sampled cases but aliasing occurs in the under-sampled scenario to misrepresent the SF of the grating to a lower value.

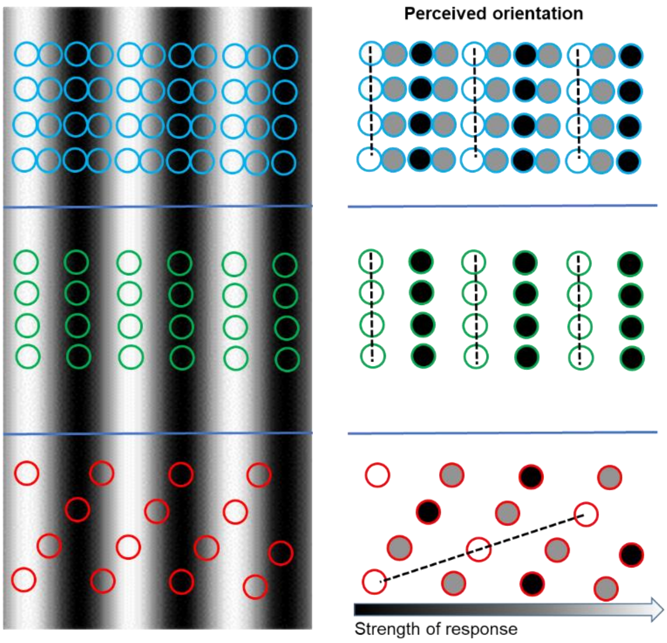


Figure 1.11: Schematic diagram demonstrating the resolution outcomes of neural over-sampling (top row), critical-sampling (middle row) and under-sampling (bottom row) *(adapted from Thibos, 1998)*. The grating is represented correctly in the over- and critical-sampled scenarios but is misrepresented in both orientation (middle column) and SF (last column) in the under-sampled case as a result of aliasing.

Aliasing is not normally perceived under foveal viewing conditions since higher SFs which cannot be veridically resolved by the neural retina are normally filtered by the optics of the eye (Campbell and Gubisch, 1966), reducing their contrast to below visual threshold. Under foveal viewing conditions, the optics of the eye are said to be the limiting factor to resolution. The sparser mosaic of photoreceptors and the higher convergence to the ganglion cells in the extra-foveal retina mean that in these locations SFs higher than the neural retina can veridically resolve, pass through and aliasing can often be perceived (Thibos et al., 1987b, Thibos, 1998). Here, the neural sampling density and not the optics of the eye are the limiting factor (Anderson, 1996b). The highest SF that can be accurately represented by the neural array is called the Nyquist frequency and according to the Shannon sampling theory (as cited in Green, 1970) this is one half of the neural sampling frequency.

Thus four significant observations support a sampling-limit to resolution:

- The first of these is that detection and resolution acuity thresholds for gratings are different under peripheral viewing, with detection significantly better than resolution acuity (Thibos et al., 1996).
- Secondly, aliasing which occurs due to neural under-sampling has been appreciated such that the grating appears to be of lower SF and of different orientation to that presented (Thibos, 1998, Thibos et al., 1996).
- Thirdly, resolution acuity is relatively robust to the effects of optical defocus (Anderson, 1996b, Wang et al., 1997) and reducing contrast (Anderson, 1996a) compared to detection acuity for gratings.
- Lastly, several studies have demonstrated a close correlation between resolution acuity and the anatomical estimates of photoreceptors in the parafovea (Williams and Coletta, 1987) and ganglion cell density beyond 10 to 15 degrees of eccentricity (Thibos et al., 1987a).

Thibos (1993) hypothesised that these findings suggest that grating resolution could be used in the non-invasive measurement of the density of retinal neurons in the living eye in order to detect any ocular disease which results in the loss or reduction of neural structure such as glaucoma. Furthermore, conditions which result in an abnormally low neural sampling density at the fovea, such as in AMD, may also convert from optical to sampling limited resolution in the fovea, evidenced by a difference in detection and resolution thresholds for stimuli with the same average luminance as the background. Being the coarsest array in the retina, it will most probably be the ganglion cells that limit resolution, at least for gratings. This has been demonstrated in highly myopic eyes (Chui et al., 2005). Whilst repeatable sampling-limited estimates of RGC density can be made in many subjects with such a method, testing using gratings turns out to be time-consuming, variable (owing to the high guess rate in a low AFC task) and poorly understood by clinical patients.

The final stage in the Recognition Pyramid is recognition, whereby the arrangement of the letter strokes leads to letter identification. This last stage is dependent on the observer's cognitive capabilities. Factors which can influence this include the alertness of the observer, prior knowledge of the alphabet set (Cappe et al., 2014) and willingness to participate. An inability to identify a letter on current conventional letter charts reveals an inadequacy at one or more levels of this pyramid sequence but as discussed by Thibos and Bradley (1993), does not allow for the examiner to ascertain at which level/s failure occurred to determine whether optical, neural or other factors are responsible. Furthermore, whilst current VA charts are sensitive to the effects of optical defocus, thus making excellent refraction targets, their sensitivity to neural deficits appears to be relatively poor.

1.6.1 Ideal observer models

Ideal observer models which have been proposed by several groups (Watson and Ahumada, 2012, Watson and Ahumada, 2015, Nestares et al., 2003) aim to combine the optical, neural and cognitive processes described above that predict visual performance in letter recognition tasks. An ideal observer is a theoretical device defined in statistical terms that performs a specific task in an optimal way, based on the available information and specified constraints. Thus the ideal observer provides an index of the achievable level of performance when all the information is optimally used to maximise performance. A comparison of the ideal observer performance to the real performance establishes whether subject performance is limited by the available stimulus information or by perceptual or cognitive information processing limitations of the subject. One of the most recent models is the Neural Image Classifier by Watson and Ahumada (2015). This is based on their previous model (Watson and Ahumada, 2012) with refinements and additional features to address variations in letter identification with letter size by accounting for differences in spatial filtering and sampling by the RGCs with eccentricity.

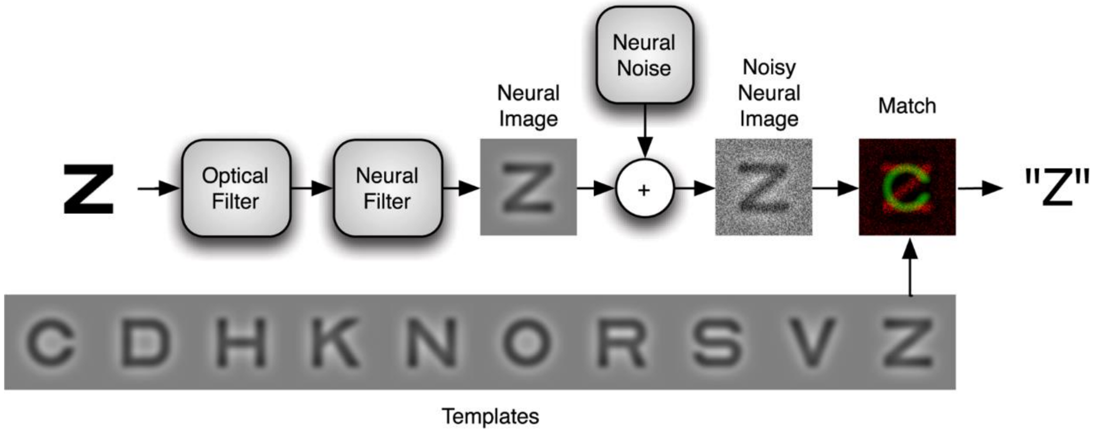


Figure 1.12: Neural Image Classifier in the identification of Sloan letters *(reproduced with permission from Watson and Ahumada, 2015. Letter identification and the Neural Image Classifier. J Vis, 15, 1-26. Copyright ARVO).*

Figure 1.12 summarises the steps which lead to the identification of Sloan letters. The letter stimulus of a particular size, presented to the observer is blurred by a filter simulating the optical filtering effects of the eye. The second stage in the model is the neural filtering of the blurred retinal image simulating the action of the receptive field of the midget RGCs. Gaussian noise representing the constant output noise of the RGCs is then added to the optically and neurally filtered image. The template (neural image of a fixed sized letter) with the highest correlation to the noisy neural image is then identified as the letter. Comparisons of predictions with this model to both historical data sets and to new measurements in three observers demonstrate that it provides good overall estimates for letter identification as a function of size (Watson and Ahumada, 2015). Another model by Beckmann and Legge (2002) underestimated the effect of eccentricity on threshold letter size, attributed to the omission of midget RGC receptive fields, sampling and noise in their model. As discussed in Section 1.5, large letters are identified using letter edges or the high SF information whereas smaller letters are identified by the gross letter strokes or low SF information. Watson and Ahumada's model complements this such that optical and retinal filtering of the same letter of different sizes produces different neural images (in which the high SF information becomes less available as the letter gets smaller) which are then compared to neural templates. The earlier version of this ideal observer model (Watson and Ahumada, 2012) was used to predict acuity performance with the Sloan letter set and other letter sets of increasing complexity. Whilst this model predicted reasonably well i.e. with high efficiency, the relative performance of the Sloan letters and the simplest letter set, it declined with increasing letter set complexity. One of the explanations given is that the neural templates with which the filtered letter is matched is imperfect and this imperfection increases with complexity. The ideal observer did correctly predict an increase in threshold size with letter complexity. This is a result of the increase in the higher SF information contained in more detailed targets and thus in order to be resolved, an increase in size would be expected.

1.7 High-pass spatial frequency letters

Conventional black-on-white letters have two distinct visual thresholds, even in the fovea; the detection and recognition thresholds. Letters can be identified using the high SF information until this becomes too small for this information to be resolved and it is filtered by the eye's optics. The visual system is able to then use the low SF information for letter identification. The difference in the low SF content between letters can help in guessing the identity of a blurred letter and indeed, certain patients can achieve superior levels of VA, owing to cognitive factors and previous visual experiences (Cappe et al., 2014). As the letters continue to get smaller still, eventually the stimulus energy in these lower SFs reaches a level whereby it becomes impossible to discriminate the letter above the chance level even though its presence can still be detected. Variability in VA measurements can often occur as a result of this transitional zone between resolution and detection (Koskin et al., 2007).

In 1978, Howland et al., (1978) designed a group of optotypes which they called 'high-pass SF letters'. These letters are designed of black and white components such that the low SFs are effectively absent. The letters are presented on a grey background so that their mean luminance equals that of the background, similar to grating acuity. As the letters become smaller, their narrow band of higher frequencies are filtered by the eye's low-pass optics. With the low SF information effectively unavailable for the visual system to use, and with no overall luminance cue, the letters appear to vanish. In short, the detection and resolution (and in this case recognition) thresholds of the high-pass SF letters are closely similar. This is the reason these letters are often referred to as 'Vanishing Optotypes' (VOs). While still containing a range of SFs in their Fourier spectra, VOs have a more band-pass design, the lower object SFs being notably absent (Figure 1.13) (Charman, 2006).

Figure 1.13: Fourier magnitude spectra for a conventional and VO letter *(reproduced from Howland et al., 1978)*. Appearance of (a) conventional black and white letter (top) and (b) Howland's VO letter (top) with their respective Fourier magnitude spectra (bottom). Low SFs are closer to the centre with increasingly higher SFs on moving away. The low SFs can be seen to be absent in the VO letter, as is the central DC. Contrasts may not be correct due to reproduction limitations.

Striped targets with the same mean luminance as their background were actually first described by Friedenwald for use in a chart to test for astigmatism (Friedenwald, 1924). We also saw in Section 1.4 how Walker (1942) designed a test made up of equal black and white squares on a grey background. Harris et al., (1984) used the 'vanishing' property of these targets in the technique of preferential looking in infants. They found steeper psychometric functions using complex face stimuli constructed in a VO design suggesting that more precise estimates of acuity could be achieved compared to more simple square wave grating stimuli. Medina and Howland (1988) created a high-pass frequency letter chart (Figure 1.14) with each limb constructed of five-element strips of black and white lines with relative widths of 1, 2, 3, 2 and 1 units. They based their chart design on a Snellen chart with letters scaled in size with a non-constant ratio to the corresponding line on the original Snellen chart based on their findings of difference in acuity between the two charts.

Figure 1.14: Medina and Howland's high-pass frequency acuity chart *(reproduced from Medina and Howland, 1988)*. Contrasts may not be correct due to reproduction limitations.

In 1986, Frisen proposed that the 'vanishing' property of these targets could be usefully incorporated into a visual field test which may be more patient-friendly than conventional perimetry tests (Frisen, 1986).

1.7.1 The Cardiff Acuity test

The most familiar use today of VOs in clinical practice, exploits their 'vanishing' property in the technique of preferential looking with a picture test (Woodhouse et al., 1992, Adoh et al., 1992, Adoh and Woodhouse, 1994, Fariza et al., 1990) in assessing acuity in toddlers and children with intellectual impairment. Dr Maggie Woodhouse introduced her Cardiff Acuity cards (Woodhouse et al., 1992) in which familiar pictures, selected from Kay's study of the recognisability of pictures in young children (Kay, 1983) are presented in high-pass format (Figure 1.15). The borders of the picture consist of a white centre flanked by a black edge on either side, each half the width of the central white part (Adoh et al., 1992, Adoh and Woodhouse, 1994, Woodhouse et al., 1992). With better acuities, the width of the borders get narrower and narrower but the overall size of the picture remains constant. The picture is presented either at the top or bottom of a card and is judged as being detected, therefore implying it has been resolved, if the infant directs their gaze toward the picture (Woodhouse et al., 1992). From a test distance of 1 m, an acuity range of 0.00 to 1.00 logMAR can be measured in 0.1 logMAR steps with three cards presented at each logMAR level.

The test was originally designed as a more interesting alternative to preferential looking techniques using gratings in order to maintain better attention in toddlers aged 1 to 3 years old (Woodhouse et al., 1992) and the clinical use of the

Cardiff acuity cards is now well established (Mackie et al., 1995, Adoh et al., 1992, Adoh and Woodhouse, 1994). Adoh and Woodhouse (1994) suggest slightly worse acuities when the task is changed from a detection (which therefore implies resolution) to a recognition one in children, and reasoned this is probably due to the increased cognitive demand that naming a picture requires.

Figure 1.15: Example of a Cardiff Acuity card *(reproduced from Charman, 2006).* The image luminance may not be balanced due to reproduction limitations. The image is presented either at the top or bottom of the card and acuity assessed using the preferential looking technique**.**

Two studies have looked at the sensitivity of the Cardiff Acuity test to detecting refractive blur. Compared to conventional letter charts, using the Bailey Lovie chart (Howard and Firth, 2006) and the ETDRS chart (Paudel et al., 2017), the Cardiff Acuity test was found to be less sensitive. Studies have demonstrated that conventional letters are particularly vulnerable to the effects of phase reversals associated with optical defocus, whereas grating acuity remains relatively unaffected due to spurious resolution (Thorn and Schwartz, 1990). Gratings are affected negatively only by the modulus of the optical transfer function (OTF) which produces a contrast loss of the grating. The phase-shift of the OTF results in reversed contrast of the grating. It may be that for the targets in the Cardiff Acuity test, a similar process happens whereby the higher frequencies at the edges phase reverse, increasing the effective stroke width of the image making it more easily detectable (Thorn and Schwartz, 1990). In contrast, complex patterns such as letters are also affected by the phase of the OTF such that phase distortions induced by the OTF cause artefacts such as contrast reversals and multiple images which, when added to the contrast losses, quickly impair the recognition task (Nestares et al., 2003, Ravikumar et al., 2010).

The Cardiff Acuity cards have surprisingly been shown to be relatively less effective at detecting amblyopia which is a neurodevelopmental disorder (Geer and Westall, 1996). The reason for this however, appears to be related to the actual test rather than stimulus design where the comparison study compares a crowded letter identification task to a simple detection/resolution task with the uncrowded Cardiff cards. Since crowded letter charts are known to be more sensitive in detecting amblyopia deficits (Levi, 2008), this finding is not surprising. Furthermore, the scoring techniques used, allowed for better sensitivity values with the conventional letter crowded tasks (Geer and Westall, 1996).

1.7.2 High-pass resolution perimetry

The use of VOs over gratings as a non-invasive measure of neural sampling density has also been incorporated in a high-pass resolution perimetry (HRP) test. Frisen proposed the detection of a single high-pass filtered closed ring to obtain an acuity related measurement in visual field testing (Frisen, 1987, Frisen, 1988, Frisen, 1993, Frisen, 1991). Conventional perimetry thresholds are determined by varying the luminance of a fixed size stimulus. HRP is available commercially as the Nikon Ophthimus Perimeter. In this test, detection thresholds of a high-pass ring stimulus of varying size but fixed contrast are used to interpret the numbers of functional neural channels (ganglion cells). Studies comparing HRP to conventional perimetry have demonstrated the benefits of HRP over conventional perimetry (Chauhan, 2000). HRP has been shown to detect the progression of established visual field loss earlier than with conventional perimetry (Chauhan et al., 1999). Furthermore, intra-test variability is unaffected by threshold or subject age unlike conventional perimetry (Chauhan and House, 1991) or eccentricity (Wall et al., 2004) and test times are comparatively shorter (Chauhan et al., 1999, Birt et al., 1998).

The targets in HRP have a light core with dark inner and outer borders (Figure 1.16). The ring width is 1/5th of the ring diameter with diameters ranging between 0.8 and 20 degrees of angle. The smallest ring is designated as ring No. 0 with a scale factor between rings of approximately 1 decibel (dB, $\log_{10}$ units or about 1.26) with 14 different ring sizes employed. The ring cores have a luminance of 25 cd/m^2, the borders 15 cd/m^2 and the background a luminance of 20 cd/m^2. The Michelson contrast calculation (Equation. 1.2), where L_{max} and L_{min}

represent the highest and lowest luminance values can be used to calculate a within-target contrast level of 25%. The reason a reduced target contrast is used is to prevent a ceiling effect limited by the resolution of the screen.

$$\text{Michelson contrast} = (L_{max} - L_{min}) / (L_{max} + L_{min})$$

(Eq. 1.2)

50 locations within the central 30 degrees of the visual field are tested in random order outside the central 5 degrees of visual angle. Stimuli are presented for 165 ms which is just below the reaction time for changing fixation.

Figure 1.16: High-pass ring stimulus used in HRP *(reproduced from Chauhan et al., 1999)*. The image luminance may not be balanced due to reproduction limitations.

The testing algorithm runs in five phases with thresholds defined as a single reversal of response for 1dB steps. Results are displayed in terms of threshold scaled ring size and the clinical parameter 'estimated neural capacity' is provided which is reported to be a measure of the underlying ganglion cell density (Frisen, 1991).

Calculations for the number of ganglion cells (N) per degree of visual angle are given by Frisen (1988) using the ganglion cell spacing in arcmins (S):

$$N = 1 / (60 \times S)$$

(Eq. 1.3)

S is calculated using the minimum angle of resolution (MAR) and F which is the proportionality factor:

$$S = MAR / F$$

(Eq. 1.4)

A 1dB increase in test threshold represents a 39% decrease in ganglion cell number, 2dB a 62% decrease and 3dB a 78% decrease according to Frisen (1988).

The important point of consideration here is that the test measures detection thresholds to the ring stimulus, making the assumption that, because of the 'vanishing' design of the stimulus, this also reflects resolution thresholds for these targets. However, as previously discussed, multiple grating studies have demonstrated that, whilst detection and resolution thresholds are similar under foveal viewing conditions, they separate under non-foveal viewing conditions with resolution and not detection representing the neural sampling density (Thibos, 1998, Thibos et al., 1987a, Wang et al., 1997, Anderson, 1996b). It has been suggested that, since VO letters contain a number of different SFs at different orientations and phases, they cannot be assumed to behave in the same way as gratings (Charman, 2006). Anderson and Ennis (1999) therefore investigated these thresholds under peripheral viewing conditions for VO letters. Their study, as subsequently have others (Demirel et al., 2012) confirm that resolution and detection thresholds outside the fovea are in fact different for these targets, as for gratings with a mean luminance the same as the background. Demirel et al., (2012), discovered that resolution acuity can be either contrast or sampling limited for high-contrast, high-pass letters and this very much depends on the actual letters chosen. For those VO letters in which resolution is sampling limited, they explain that, as found for gratings, the detection acuity is set by the size of the receptor fields whilst resolution acuity is determined by the spacing (Demirel et al., 2012). Thus their designation as 'vanishing' optotypes is somewhat of a misnomer outside the fovea. HRP however uses targets of much lower contrast, background luminance and number of cycles, all of which are shown to have a detrimental effect to detection performance. Ennis and Johnson (2002) sought to investigate if in HRP, detection and resolution thresholds for the ring stimulus may once again coincide even under non-foveal viewing conditions due to the test design parameters. Their study demonstrated that whilst the detection and resolution thresholds are similar under these conditions, this is the result of optical and not sampling limited reasons since increasing target contrast improved detection performance.

1.8 Research aims

This introduction has highlighted the importance of VA testing in the assessment of visual function (Section 1.1). We have seen the significant historical developments that have taken place, particularly in the 19th and 20th centuries in terms of test design, scoring and termination criteria and letter and optotype choices in order to develop acuity tests which provide accurate and reliable results (Section 1.2). Section 1.3 demonstrated how current gold standard acuity tests still fail on two levels. Firstly they often do not give a true reflection of the status of the visual system. Whilst conventional letter charts are good at detecting optical defocus and therefore make excellent refraction targets, they often do not detect neurological causes of loss of vision accurately. Secondly, test variability can often hinder the detection of any changes. Section 1.4 discussed how one source of this variability may arise from the different legibility values of letters. Letters particularly vary in their low SF content identified in Section 1.5. For some letters, this is more similar than others. It therefore may be easier to correctly guess the identity of certain letters when they are blurred and this ability can vary in patients depending on cognitive ability and previous experience. Section 1.6 discussed the optical, neural and cognitive limits to visual acuity and how ideal observer models have been used to represent these. Section 1.7 introduced us to the unique properties of high-contrast, isoluminant targets where resolution thresholds can be used to indicate the neural sampling density of the visual system.

Whilst high-pass filtered targets are currently used clinically in grating, picture and ring format, the primary aim of this research was to investigate and design a test/s using these targets in letter format that would better allow for the non-invasive measurement of localised retinal function in AMD. Letter targets are preferable since their wide use in visual function tests make them familiar and the large nAFC available allow for a low guess rate. The advantage in using high-pass filtered versions of letter targets for the purposes of measuring visual function in AMD are investigated and may address some of the issues related to accuracy and repeatability of current conventional tests.

In order to achieve this, a number of initial laboratory based studies were required to further investigate and explore the properties, characteristics and behaviour of VO letter stimuli both under foveal and peripheral viewing conditions. These investigations led to the creation of the Moorfields Acuity chart; a logMAR based VA chart incorporating high-pass filtered letters. This was tested in clinical studies to investigate localised retinal function in normal subjects and in patients with AMD.

1.8.1 Laboratory based studies

Aim 1: To determine the effect of the nAFC on VA measurements and variability using VO letters compared to conventional letters (Chapter 3).

Aim 2: To determine relative recognition thresholds of VO letters across the whole alphabet in order to investigate if between-letter performance is indeed less variable than for conventional letters (Chapter 4).

Aim 3: To measure the detection and recognition performance for each individual letter of the high-pass alphabet in foveal and peripheral vision in order to determine which letters display sampling limited performance (Chapter 4).

Aim 4: To investigate the effect of optical defocus on the detection and recognition thresholds of high-pass letters in the fovea and periphery (Chapter 4).

Aim 5: To investigate the effect of simulated lens ageing on recognition thresholds of high-pass letters in the fovea and the periphery (Chapter 5).

1.8.2 Clinical studies

Aim 6: To design a clinical logMAR style VA test using carefully selected high-pass letters and determine how threshold measurements and repeatability vary with test termination criteria and scoring methods in normal subjects with uncorrected refractive error (Chapter 6).

Aim 7: To determine how VA thresholds with the high-pass logMAR chart compare to those using a conventional letter chart in patients with AMD (Chapter 7).

Aim 8: To investigate detection and recognition thresholds for high-pass letters in AMD patients (Chapter 8).

2. Methodology

CHAPTER 2

This chapter provides a description of the equipment and methods for data acquisition. The preliminary laboratory studies (Aims 1 to 5) were conducted using electronically generated and presented stimuli. The clinical studies used printed hard-copy VA charts (Aims 6 and 7) and electronic stimulus presentation (Aim 8). A summary of the statistical tests used for analysis is also provided. Further detailed information is given in the Methods section of the relevant chapters where appropriate.

2.1 Computer based studies

2.1.1 CRT monitors

Cathode ray tube (CRT) monitors have traditionally been considered to be the gold standard monitor for use in the psychophysical and physiological studies of vision. The image on a CRT monitor is generated by the excitation of a phosphor layer which emits light when struck by an electron beam. The electron beam scans from the top left to the bottom right corner of the monitor. The decay in the fluorescence of the phosphor in combination with the scanning of the electron beam results in a flicker at 60 to 120 hertz. Whilst invisible to the eye, this can make exact calculation of the stimulus duration difficult (Mulholland et al., 2015). There is also a lack of independence between neighbouring pixels such that interactions between adjacent pixels can affect luminance values (Pelli, 1997). CRTs are no longer commercially available, having been superseded by lighter and flatter liquid crystal display (LCD) monitors. However, whilst the limitations of CRT monitors are recognised, the critical shortcomings of current LCD displays mean that these screens are not widely used in vision research.

LCD screens consist of a layer of liquid crystals sandwiched between two polarising filters with a light source at the back of the monitor. As voltage is applied to the layer of liquid crystals, they align to block the light from passing through, the intensity of which is controlled by the voltage applied. Whilst these monitors do demonstrate pixel independence, the sluggish change induced in the liquid crystals means they are slower in response compared to CRT monitors with greater variability in response times over repeated measurements (Elze and Tanner, 2012). The resulting image persistence causes problems such as motion blur in rapidly changing stimuli (although some newer models of LCD screen overcome this using compensatory image processing) compared to the shorter-persistence phosphors used in CRT monitors (Wang and Nikolic, 2011, Ghodrati et al., 2015). In addition, the luminance of LCDs can depend strongly on the viewing angle, a problem that persists even with newer models (Wang and Nikolic, 2011, Ghodrati et al., 2015). This is particularly relevant for experiments reported in this thesis in which stimuli are presented in off-centre locations whilst the observer maintains central fixation.

The computer-based studies for this research were thus conducted on one of two gamma (γ) corrected high resolution (1280 x 1024 pixels) CRT monitors; Dell Ultrascan P991 CRT monitor (Dell Corp. Ltd, Brackness, Berkshire, UK) for Aim 1 and Dell Trinitron P992 CRT monitor (Dell Corp. Ltd, Brackness, Berkshire, UK) for Aims 2 - 5 and 8. A warming up period for the screens of at least half an hour was allowed before any experiments were conducted in order to ensure that stable luminance outputs were achieved.

2.1.2 Visual stimuli

The high-pass stimuli (Figure 2.1) and conventional letters, both of a 5 x 5 unit grid design, were generated using MATLAB (version 7.6, Mathworks, Inc., Natick, MA, USA) and were presented using an Apple Macintosh computer (Apple, Inc., Cupertino, CA). True 14-bit contrast resolution was achieved in hardware using a Bits++ video processor (Cambridge Research Systems, Ltd., Rochester, UK).

Figure 2.1: The high-pass filtered letter set

The VO stimuli of a pseudo high-pass design were constructed with an inner black core flanked by a white border half the width of the central section. These stimuli were presented on a grey background such that the mean luminance of the letter was the same as the background. A 'stroke width' in these stimuli was considered to be the dark centre with the two white flanks. Screen luminance was measured using the Konica Minolta Chroma Luminance Meter (CS-100, Tokyo, Japan).

2.1.3 Experimental methods

In psychophysical experiments, four main methods which are described below are employed for determining end point thresholds. For the studies reported in this research, the method of limits or adjustment and maximum likelihood adaptive procedures were used.

Method of limits (MOL) or adjustment – The subject is asked to control the level of the property of interest of the stimulus ('stimulus level') until a particular criterion is satisfied.

Staircase procedures – The stimulus level starts at a high level and is reduced until a mistake is made by the subject at which point the staircase may reverse and is increased until a correct response is attained, whereupon another reversal takes place and so forth. The values of the reversals are typically averaged. The behaviour of the staircase can be altered using different step sizes and termination rules.

Maximum likelihood adaptive procedures – These are staircase procedures in which the next presentation level of the stimulus is positioned each time at the maximum likelihood estimate of threshold using available information; specifically the subject's previous responses and prior knowledge of their likely performance. QUEST (Watson and Pelli, 1983), an example of such an adaptive staircase procedure uses prior knowledge (represented by the prior probability density function) combined with information from previous trials (represented by the likelihood function) to form the posterior probability density function, to decide the placement of the subsequent trial. The final estimate of threshold is based however only on the modal value of the likelihood function, in order to avoid bias induced by prior experience. QUEST is used in a number of studies in this thesis with termination rules specified as a fixed number of trials.

Method of constant stimuli – The stimulus level is presented at a range of predetermined values in an interleaved random order which prevents prediction or expectation but can result in a large number of trials. The percent correct for each level is then calculated and a frequency-of-seeing curve plotted.

2.1.4 Stimulus presentation times

Seiple et al., (2001) demonstrated that thresholds for letter identification require a stimulus duration far longer than that needed for letter detection. The computer based test letter stimuli in the studies in this thesis were presented for 500 ms, based on the reported critical duration time for foveal visual acuity based on letter identification for background luminances ranging from 10 to 200 cd/m^2 (Niwa and Tokoro, 1997). Whilst critical duration times for letter identification are longer in the periphery than the fovea (Seiple et al., 2001), stimuli durations of 500 ms were also used for peripheral resolution experiments in order to prevent thresholds being affected by eye movements (voluntary or involuntary) to the stimulus location.

2.1.5 Refractive error determination and correction

Refractive error for foveal viewing was determined initially using retinoscopy. The appropriate lenses were placed in a trial frame in front of the test eye in line with the foveal stimulus with the fellow eye occluded. This was subjectively refined by maximising the contrast of a high SF target displayed on the CRT monitor at the relevant working distance (Thibos et al., 1996). A similar procedure was used for refractive error correction at 10 degree eccentricity in the nasal field (temporal retina) used for a number of the studies with the lenses positioned in front of the test eye in line with the peripheral stimulus.

2.2 Visual acuity chart studies

2.2.1 Visual acuity chart production

Both the Moorfields Acuity Chart (MAC) and conventional black-on-white logMAR charts (described in detail in Chapter 6) were graphically designed and digitally produced by the reprographics department at Ulster University, Coleraine, N. Ireland.

2.2.2 Chart luminance

Chart luminance was measured using the Konica Minolta Chroma Luminance Meter (CS-100, Tokyo Japan). Three different VA chart types were used in the studies. In line with the recommendations of the ICO, the minimum background luminance of the internally illuminated ETDRS charts used for refraction were confirmed to be no lower than 160 cd/m^2 (International Council of Ophthalmology, 1988). Testing with the non-illuminated conventional black-on-white VA charts was done under normal room lighting conditions, ensuring that, as per the recommendations of Sheedy et al., (1984), the chart luminances were in the range 80-320 cd/m^2 and in line with the recommendations of the BSI, the chart luminance was more than 120 cd/m^2 with variance across the chart not exceeding 20% (British Standards Institute, 2003). Testing with the MAC charts was conducted under these same room lighting conditions.

2.2.3 Chart viewing times

As per normal routine clinical practice, the viewing time with the VA charts was unrestricted. Subjects were not allowed to change their letter identity response once a subsequent letter had been attempted.

2.2.4 Refractive error determination and correction

Study measurements were taken on clinical subjects using the appropriate refraction results unless otherwise specified. Monocular refraction in negative cylindrical format was established using the ETDRS chart R at a distance of 4 m (or 1 m in cases of poor VA described in Section 2.2.5) with retinoscopy and conventional subjective refraction procedures.

2.2.5 Visual acuity score calculation

VA scores were recorded on predesigned proforma with a circle placed around each letter read correctly and a line through each incorrect letter. A termination rule of 5 out of 5 letters wrong on a line was used so that the effect of using different termination and scoring techniques on VA scores could retrospectively be applied and investigated.

Interpolated single-letter logMAR scores with a 4 m test distance were calculated using Equation 2.1, where VA_{4m} means VA using a 4 m test distance and C_N is the number of correct letters.

$$VA_{4m} \text{ (logMAR)} = \text{(Value of top line on chart + 0.1)} - C_N * 0.02$$

***(*Eq 2.1)**

Whilst a 4 m test distance was conventionally used, a 1 m test distance was used in cases of poor vision whereby either the observer or the test chart was moved closer. This was only applicable to the initial VA and refraction assessment process since eyes with VA beyond the range available on the MACs were not included in the study test procedures to avoid complications in comparing measurements attained at different test distances. Different protocols have different advice on when a shorter test distance should be employed with some suggesting this should be when fewer than 4 letters are read correctly from 4 m (Patel et al., 2008, Laidlaw et al., 2008) and others when less than 20 letters are read correctly from 4 m. In the studies in this thesis, when any mistakes were made on the top line with a 4 m test distance, a 1 m test distance was employed. In this case, the calculated logMAR score for a final test distance of 1 m (VA_{1m}) is adjusted appropriately by adding 0.60 logMAR, ensuring that the total number of correctly read letters at 4 m and at 1 m are taken into account (Equation 2.2).

$$VA_{1m} \text{ (logMAR)} = \text{(Value of top line on chart + 0.1)} - C_N * 0.02 + 0.60$$

***(*Eq.2.2)**

In recent times, recording acuity scores in 'logMAR' has shifted to reporting in terms of 'number of ETDRS letters'. This is a more intuitive measure for patients to understand the number of letters gain/loss in their VA. The calculation of the score in terms of number of ETDRS letters with a 4 m test distance is given by Equation 2.3:

$$VA_{4m} \text{ (number of ETDRS letters)} = (1.6\text{-top line of chart in logMAR}) / 0.02 + C_N$$

(Eq. 2.3)

For a 1 m test distance, the score is simply the number of letters read correctly (Equation 2.4):

$$VA_{1m} \text{ (number of ETDRS letters)} = C_N$$

(Eq. 2.4)

2.3 Forced choice procedures

Subjects differ in their willingness to guess the identity of a letter when close to threshold. In order to reduce any effects that result from differences in response criteria (Section 1.3.3), a forced choice test procedure was employed for all clinical and laboratory based studies. Each subject was asked to read each letter until the pre-set termination criterion was fulfilled. When unsure of a letter identity, the subject was encouraged to guess since this has been demonstrated to improve VA scores (Smith, 2005).

2.4 Test randomisation

Using the ETDRS charts, studies report a mean improvement in VA with repeated measurement, which can be attributed to learning. Whilst this effect is small ranging from no effect to an improvement of 1.5 letters (Ferris et al., 1982, Laidlaw et al., 2008, Shah et al., 2011b), this was minimised by carrying out all test measurements in a random order. This was also done to control for fatigue since each individual underwent numerous VA measurements. The random test order was generated using an online pseudo-random number generator (http://www.randomizer.org/).

2.5 Statistical analysis

All statistical analyses were conducted using the GraphPad Prism statistical analysis package (version 5.04, GraphPad Software, La Jolla, California, USA). This section explains and summarises the main statistical tests used in the data examination for the studies in this thesis.

2.5.1 Testing for a normal distribution

In parametric testing, the underlying assumption is that the data set was sampled from a normal or Gaussian distribution. This can be verified graphically by constructing a frequency distribution and looking for the absence of a positive or negative skew. This can also be tested numerically using a test of normality. In this thesis, the Shapiro-Wilk W test on the GraphPad statistical software was employed since it is appropriate for small sample sizes (less than 50) but is also capable of handling sample sizes as large as 2000 and is widely considered to be the most powerful of the normality tests (Razali and Wah, 2011). The Shapiro-Wilk W test uses the null hypothesis that data are sampled from a normal distribution. If the p-value is less than 0.05, the null hypothesis is rejected and the conclusion reached that there is evidence that the sample data are not from a normally distributed population. Conversely, if the p-value is greater than 0.05, the null hypothesis cannot be rejected and the conclusion is drawn that any deviation of the sample data from a normally distributed population is not more than would be expected from chance alone. Statistical tests tend to be quite robust to violations of the normality assumption, especially with sample sizes of more than 30 so that it may still be appropriate to use parametric statistical tests in these instances (Ghasemi and Zahediasl, 2012).

2.5.2 Bland-Altman analysis

Bland-Altman analysis is preferable over correlation and linear regression analysis in describing the agreement between two quantitative measurements (Bland and Altman, 1986). Here, Bland-Altman analysis was used to compare VA measurements attained with logMAR design charts using conventional black-on-white letters (considered as the gold standard) compared to measures made with charts using high-pass filtered letter designs.

Correlation can reveal whether and how strongly pairs of variables are related with the correlation coefficient, r, measuring the strength and direction of linear relationship between two variables. Linear regression analysis finds the line that best predicts one variable from the other with the coefficient of determination, r^2, indicating how well the data is represented by the regression line. However, neither of these gives information on the level of *agreement* between the two measurements (Bland and Altman, 1986, McAlinden et al., 2011, Giavarina, 2015).

In Bland-Altman analysis, LOA are calculated using the mean and SD of the differences between the two measurements to describe by how much they differ. A scatter plot can be created in which the average of the two paired measurements (in this case, VA measures from the two different chart types for each subject) on the x-axis is plotted against the difference between measures on the y-axis.

The agreement is described by the mean bias or difference (d) with 95% of the differences expected to lie within d +/-1.96SD of the differences if these follow a normal distribution (see Section 2.5.1). These upper and lower limits are termed the 95% LOA. A systematic bias is considered significant if the line of equality is not within the confidence interval of the mean difference. An upward or downward trend in the scatter plot indicates a proportional error suggesting that the mean difference between the two varies with the magnitude of the measurement.

Bland-Altman analysis can also be used to assess the precision of a test by taking repeated measurements with that test. The mean bias would now be expected to be approximately zero since both measurements are taken using the same test so that any difference would be attributable to a learning effect. Once again, if the differences are normally distributed, 95% of the differences between the test and retest would be expected to lie within d +/-1.96SD of the dataset, which represents the coefficient of repeatability or TRV. As mentioned in Section 1.3.3, TRV values of up to 2 logMAR lines have been reported with logMAR charts. An increase or decrease in the level of scatter suggests that the variability of the test is dependent on the magnitude of the measurement.

2.5.3 Ordinary least-squares linear regression

In instances where Bland-Altman analysis suggested a proportional bias in addition to a systematic bias, ordinary least-squares linear regression analysis was performed to investigate the relationship between the independent or predictor variable (in this case mean value of VA on the x-axis) and dependent or criterion variable (in this case difference in VA between charts on the y-axis). With one independent variable, this simple relationship is established using a best fit straight line, the regression line, through the points. The equation for the regression line is $y = mx + c$ whereby the y-term is predicted using the x-term, the gradient of the slope (m) and the y-axis intercept value (c). The vertical line from each point to the regression line represents the errors of prediction (or residuals). The most common method employed for fitting the regression line is the ordinary least-squares method whereby the sum of the squares of the residuals are minimised. The goodness of fit of the regression line to the data is quantified by r^2 ranging from 0 to 1. A reported value of 0 implies that knowing the x-value does not help in predicting the y-value, whilst conversely, a value of 1 allows the y-value to be predicted perfectly.

2.5.4 Comparing mean values between data sets

The Student's t-test was used to assess whether the means of two groups are statistically different from each other relative to the variability of their values. The null hypothesis was set that no difference exists between the means in question, with any discrepancy being attributed to chance and the student's t-test will tell us if the data are consistent with this. The calculated t-value is compared to a tabulated value and in this thesis the 95% confidence level is considered acceptable. If the calculated value exceeds this tabulated value, then the null hypothesis is rejected and a significant difference at the 95% level between the means is reported. The unmatched t-test compares the means of two unmatched groups and assumes that the two samples come from populations which follow a normal distribution with the same variance. The paired t-test is used when the two data samples are related. A one-tailed t-test is used when the difference is expected to only be in one direction in which case the null hypothesis would set the discrepancy for a certain direction either smaller or larger, otherwise a two-tailed t-test is used.

When a statistical test is carried out, there is a chance of making a type I or type II error. A type I error is such that a true null hypothesis is incorrectly rejected and an effect or difference is found that is not actually present. A type II error is the failure to reject a false null hypothesis, and the probability of making a type II error (beta level), is related to the power of the test. The probability of making a type I error is called the alpha level and this is often set to 0.05 implying that there is a 5% probability of incorrectly rejecting the null hypothesis. When a comparison of more than two sets of data are to be made, the type 1 error rate increases. An Analysis of Variance (ANOVA) test allows for control of this so that the type I error remains at 5%. A repeated-measures ANOVA allows any overall differences between related means to be investigated, the null hypothesis in this case being that the means for each data set are all equal. With a statistically significant result, post-hoc statistical tests can be applied to investigate where these differences occur.

2.5.5 Comparing variances between data sets

The F-test was used to establish whether the variances (which is the square of the SD) of two populations are equal, the null hypothesis in this case being that the variances are equal. This test, like the t-test can be one-tailed looking for a difference in only one direction i.e. either larger or smaller which would be reflected in the null hypothesis but not both, and the two-tailed test tests for a difference in either direction.

2.5.6 Sample size calculations

The recommendations of Bland and Altman (1999) were used in order to calculate a sample size for the method comparison studies between ETDRS and MAC VA measurements in a group of subjects, based on the desired confidence interval for the 95% LOA (Bland, 2004). Their calculation (Equation 2.5) provides an estimate of the confidence interval (CI) as:

$$CI = +/-1.96 \sqrt{SD(3/n)}$$

(Eq. 2.5)

where n is the sample size and SD is the standard deviation of the difference between measurements by the two VA charts.

Thus rearranging Equation 2.5,

$$n = (3 \times 1.96^2 SD^2) / CI^2$$

(Eq. 2.6)

With no previous data available for the SD of VA measurements with the MAC chart, calculations were based on reported SD values for the ETDRS chart. In an effort to make a cautious estimate of sample size, the greatest reported SD value for ETDRS measurements repeated on the same day was used. This value is 0.094 logMAR for subjects with AMD (Blackhurst and Maguire, 1989). A calculation for a CI of one letter would result in a value of n of 255 subjects. For a CI of two letters, n would equal 64 subjects.

In view of this, a minimum of 64 participants were recruited for the method comparison studies in this thesis. A larger number was recruited where possible and practical in order to ensure a reasonable spread in VA values. This seems a reasonable number when compared to other method comparison studies, comparing different VA charts to the ETDRS charts and employing Bland-Altman statistics, which have used between 40 and 50 participants (Lim et al., 2010, Rosser et al., 2001, Laidlaw et al., 2003, Bokinni et al., 2015).

3. Effect of the number of alternatives on Vanishing Optoptye acuity thresholds and repeatability

CHAPTER 3

The work discussed within this chapter has been published in a peer reviewed journal;

Shah N., Dakin S. C., Redmond T. & Anderson R. S. Vanishing Optotype acuity: repeatability and effect of the number of alternatives. Ophthalmic and Physiological Optics. 2011;31(1):17-22.

Publication acknowledgement is given to the journal Ophthalmic and Physiological Optics, the College of Optometrists and Blackwell Publishing.

3.1 Introduction

Whilst several tests assess structural changes to support monitoring of disease progression and the efficacy of therapy, VA measurement remains the universal primary test in evaluating functional change. As discussed in Section 1.1, in order to reliably determine if either a significant adverse or positive change in performance has resulted from either abnormality or treatment, any VA test should provide accurate and repeatable measurements.

As seen in Figure 1.13, conventional letters contain a range of SFs and the visual system shifts to rely on the low SF content for conventional letter recognition when the higher SFs become unresolvable as the letters get smaller in both foveal and peripheral vision (Alexander et al., 1994, Anderson and Thibos, 1999a, Anderson and Thibos, 2004, Bondarko and Danilova, 1997, Chung et al., 2002, Gervais et al., 1984, Parish and Sperling, 1991). Since letters particularly vary in their low object SF content (Anderson and Thibos, 1999a, Gervais et al., 1984, Anderson and Thibos, 1999b), it may be easier to correctly guess the letter identity between two letters which are very different compared to two letters which are more similar in this content. The work of Banister (1927) described in Section 1.4 suggests that discriminability is not the inherent property of an individual letter demonstrated by the conflicting classification of the letter 'H' as being one of the easiest, intermediate and hardest letters to identify by different research groups (Banister, 1927). Thus VA thresholds will vary depending on the probability of correctly identifying a given optotype from any number of other alternatives available and what these alternative letters choices in fact are. Indeed Carkeet (2001) used exact calculation and Monte Carlo simulation techniques to demonstrate significant differences in the mean and SD of logMAR scores as a consequence of termination criteria with different nAFC. With an increasing number of alternatives, the mean increases owing to the greater uncertainty in letter identity arising from both the increasing alternative letter choices with which the presented letter could potentially be confused and the greater likelihood of any letter having a more closely similar rival. This greater letter uncertainty with an increasing nAFC surprisingly does not actually lead to greater threshold variability, in fact possibly the opposite, since the subject is less likely to guess a letter correctly when it is unresolvable. A naïve subject may assume the presence of 26 alternatives when looking at an acuity chart with actually only 10 alternatives, whereas a more observant subject may realise the chart has a more limited number of choices. If it is possible to reduce the effect on acuity thresholds of the number of alternatives, this would be advantageous when considering different test chart designs in order to reduce variability in thresholds with termination criteria.

In a letter identification task, removing the lower SFs where conventional letters differ substantially (Anderson and Thibos, 1999a, Gervais et al., 1984, Anderson and Thibos, 1999b), would force the visual system to rely solely on available high SF content. Thus letters may become more equally discriminable and more closely similar resulting in VA thresholds which are less affected by the nAFC available. The resulting higher level of uncertainty may reduce the subsequent variation in VA scores between charts employing differing numbers of alternatives. Figure 3.1 (a) and (b) illustrate this concept where two conventional letters very different in their low SF content ('A' and 'U') will remain discriminable down to a smaller size than two letters more similar in their low SF content ('O' and 'Q'). However if the low SF content is removed as in (c) and (d), the letters may become much more equally discriminable and thus acuity thresholds may become less affected by the number of alternatives and what these actually are.

The aim of this study was to test this hypothesis by comparing the VA thresholds attained with different nAFC using pseudo high-pass filtered letters relative to conventional letters and to compare the variability in VA measurements between the two letter types.

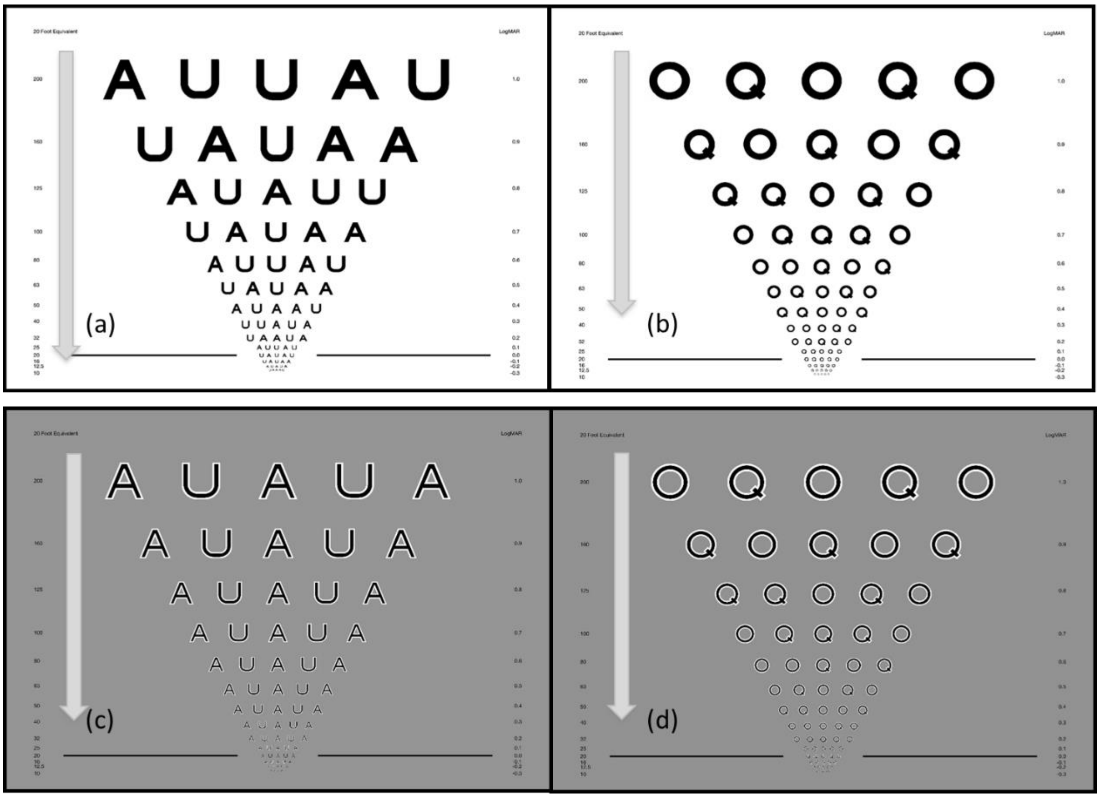


Figure 3.1: Effect of low SFs on recognition thresholds. In conventional letter format, (a) two letters that are very different in their low SF content such as 'A' and 'U' will remain discriminable down to smaller sizes than (b) two letters more similar in their low SF content such as 'O' and 'Q'. If the low SF content is removed to create high-pass filtered letters (c and d), the letters may become more equally discriminable and thresholds more similar regardless of letter choice.

3.2 Methods

3.2.1 Subjects

Three healthy experienced psychophysical observers (NS aged 27 years, RSA aged 45 years and TR aged 26 years) with no ocular abnormalities were recruited for this study. All three subjects had VA of 6/5 or better with refractive error corrected for foveal testing using the methods described in Section 2.1.5. The right eye of all subjects was tested, with subjects NS and RSA both being emmetropic whilst TR had a mean spherical refractive error of -3.00D. Ethical approval was obtained for this study from the relevant UCL research ethics committee and all procedures adhered to the tenets of the Declaration of Helsinki. All test examinations were conducted in the Spatial Vision Laboratory at the UCL Institute of Ophthalmology, London, UK.

3.2.2 Procedure

VO letters (described in Section 2.1.2) and conventional letters of a 5 x 5 unit grid design were generated using MATLAB (version 7.6, Mathworks, Inc., Natick, MA, USA) and were presented using an Apple Macintosh computer (Apple, Inc., Cupertino, CA) on a γ-corrected high resolution (1280 x 1024 pixels) Dell Ultrascan P991 CRT monitor (Dell Corp. Ltd, Brackness, Berkshire, UK). The OpenGL capabilities of the computer's built-in graphics card (ATI Radeon X1600; AMD, Sunnyvale, CA, USA) was used to achieve scaling of the stimuli. This bilinear interpolation procedure allowed the display of stimuli of arbitrary size with sub-pixel resolution while retaining the accurate representation of their balanced luminance structure. The luminance of the background of the CRT monitor was 53.9 cd/m^2 and all testing was conducted under low room illumination to avoid screen reflections. Stimuli were presented for 500 ms at a test distance of 3.8 m at which the screen subtended 5.3° x 4° with one pixel subtending 0.25 arcmins. Threshold recognition VA was determined for each subject for both conventional and VOs for different numbers of AFC. These alternative choices were 2AFC (letters AU and OQ), 4AFC (AQUO), 6AFC (QUANGO) or 26 AFC (whole alphabet) as displayed in Figure 3.2. These letter choices were based on the investigations of Anderson and Thibos (2004) in which they compared how psychophysically attained threshold acuities varied with an index for dissimilarity. The value for this index was determined by using computer analysis to subtract the image of one letter from its pair and performing a Fast Fourier Transform on the difference image to produce a difference spectrum in the frequency domain. The letter combinations 'A versus U' were found to be the most dissimilar whilst 'O versus Q' were the least.

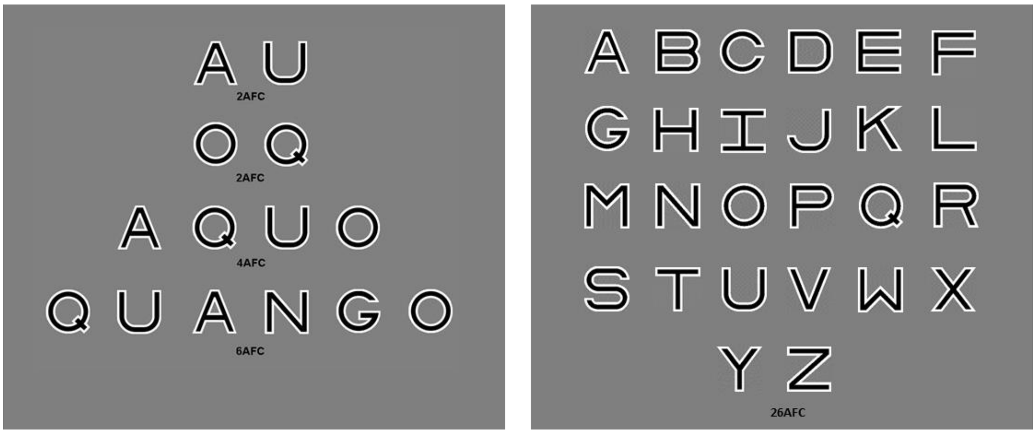


Figure 3.2: The 2, 4, 6 and 26 alternative forced choice VO letter sets *(adapted from Shah et al., 2011a)*.

Each nAFC was tested in a randomised order with the acuity threshold determined using QUEST in which the size of the letter displayed was determined by knowledge of the previous responses, with trials spread evenly on a decimal/log axis. The initial letter size displayed was 115.8 x 115.8 arcmins and the prior density function was limited by the minimum and maximum displayable letter size on the screen. The slope (β) of the psychometric function used was set to 3.5, a value widely used in psychophysical literature. Each test run involved 50 letter presentations in total with the final acuity threshold determined by QUEST's built in maximum likelihood estimation procedure of threshold. Subjects were made aware of the letter set available for each test which were also displayed in the corner of the screen. The subject's verbal response of the letter identity was entered by the examiner on the keyboard. This process was repeated until three repeat measures were attained for each nAFC.

In the QUEST psychophysical procedure, the gamma value (guess rate) was fixed at 50% with the threshold performance level (pThreshold) set at 82%. The pThreshold value determines the performance level which the psychophysical procedure converges to. For example, if set at 50% then the recognition threshold is taken when performance reaches a level at which the letter is correctly identified half the time only. Little is known about the patient's experience with different performance levels and the effect on the final thresholds that this may have. It could be possible that motivation may be better when functioning at a higher performance level where the subject may feel more confident in their responses. Furthermore, the actual guess rate that an observer may adopt is unclear. Some subjects may assume a 26 alternative choice whilst others may realise that only a select number of letters are used in the VA test. Arditi and Cagenello (1993) explain that the indeterminacy of the guessing rate reduces the test reliability and makes within-subject and inter-subject test-retest comparisons difficult.

In order to assess the effects on VA thresholds and variability of altering these parameters, repeat measures were taken as above for one experienced observer (NS) using:

a) Gamma values 50% and pThreshold 82% for each nAFC.

b) Adjusted gamma values depending on the nAFC, calculated as 1/nAFC (50% for 2AFC, 25% for 4AFC, 16.67% for 6AFC and 3.85% for 26AFC) with a fixed pThreshold value of 82%.

c) Adjusted gamma values (as above) and adjusted pThreshold values depending on the nAFC, taken as the 50% performance level between 100% and the value of the guess rate (75% for 2AFC, 62.5% for 4AFC, 58.33% for 6AFC and 51.92% for 26AFC).

3.2.3 Statistical analysis

The final threshold letter size, generated as a percentage by reference to a box size of 512 pixels, was converted to a logMAR score for further analysis where the stroke width for the VO included both the central dark bar and the surrounding white flanks as described in Section 2.1.2. The GraphPad Prism statistical analysis package (GraphPad Software, Inc., La Jolla, CA) was used to compare VA thresholds using a one-way repeated measures ANOVA and statistically significant results were investigated using Tukey's post-hoc analysis for all pairwise comparisons and Bonferroni's Multiple Comparison post-hoc analysis for selected pairwise comparisons.

3.3 Results

The mean of the three repeat threshold measurements for each subject for each nAFC was plotted in Figure 3.3 with conventional letters on the top and VO letters at the bottom. Thresholds are given in logMAR values and the error bars represent the SD of the three repeat measures.

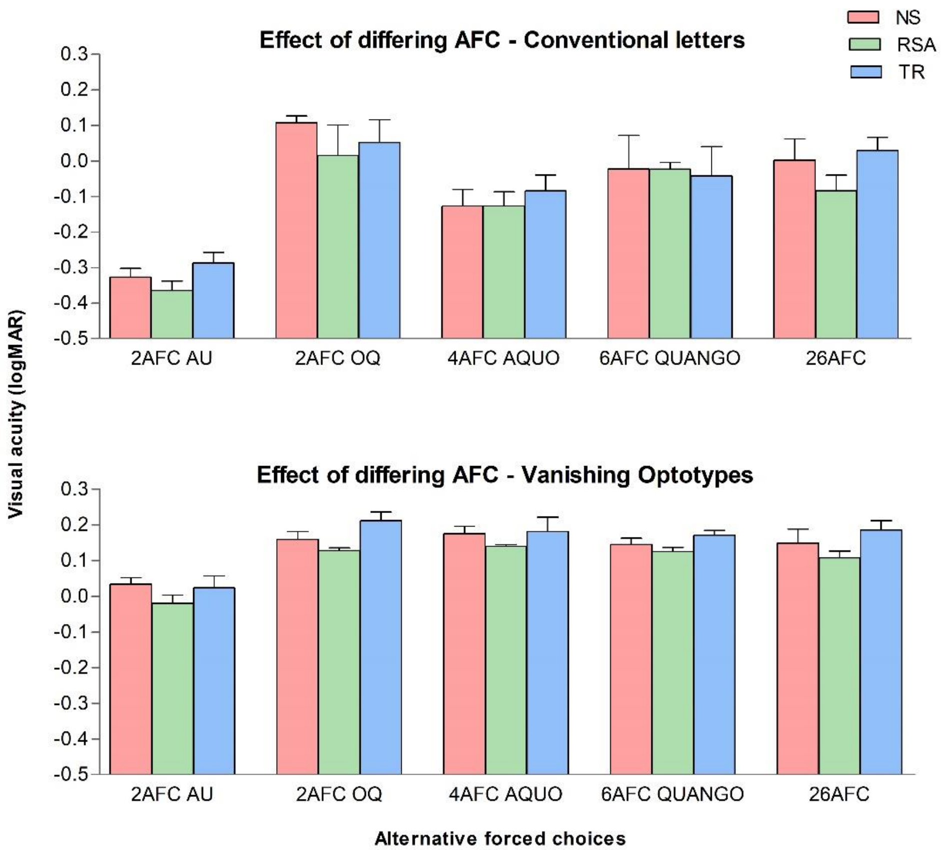


Figure 3.3: Acuity thresholds in logMAR values for each subject under each nAFC condition *(adapted from Shah et al., 2011a).* The top chart shows thresholds for conventional letters and the bottom for VOs. Error bars represent the SD of the three repeat measurements.

Firstly, it can be seen from the graphs that for every nAFC condition, the threshold letter size attained with the VOs was larger than the corresponding acuity attained with the conventional letters. A one-way repeated measures ANOVA, $F_{9,2} = 125.99$, $p < 0.0001$ confirmed this reached statistical significance and closer examination with a post-hoc Bonferroni test confirmed this held true for every nAFC at $p < 0.05$. The actual difference in acuity thresholds between the two stimulus types ranged from 0.11 to 0.34 logMAR and was not only dependent on the number of alternatives but also on what the actual letter choices were. Interestingly, it was under the 2AFC conditions that the smallest and largest between-optotype difference occurred for the letters OQ and AU respectively.

For the conventional letters, the smallest and largest mean threshold acuity values were attained in the two 2AFC conditions ranging from -0.33 (AU) to 0.06 logMAR (OQ), a difference of 0.39 logMAR. A one-way repeated measures ANOVA on acuity thresholds obtained with the conventional optotypes yielded significant differences between nAFC, $F_{4,2} = 61.72$, $p < 0.0001$ with a post-hoc Tukey test confirming significant differences at $p < 0.05$ between AU and all other AFC combinations. Significant differences were also found between OQ and AQUO, and AQUO and 26AFC (all $p < 0.05$).
The recognition acuity thresholds for the VOs were less affected than the conventional letters by the nAFC and the individual letters employed. The smallest and largest discrimination thresholds were again attained in the two 2AFC conditions but ranged only from 0.01 (AU) to 0.17 (OQ), a difference of 0.16 logMAR. A one-way repeated measures ANOVA did reveal significant differences between nAFC, $F_{4,2} = 63.13$, $p < 0.0001$ with a post-hoc Tukey test again

confirming significant differences at $p < 0.05$ between AU and all other AFC combinations, but not between any other AFC combinations. Thus, the effect of the differing nAFC is less overall for the VOs compared to the conventional letters.

Figure 3.4 shows a plot of the mean standard error of the logMAR thresholds (averaged for the three repeat readings and across subjects) for each of the nAFC conditions in the two letter designs. It can be appreciated that the variability was lower for the VOs (0.01 – 0.02 logMAR) compared to the conventional letters (0.02 – 0.04 logMAR). This is also evident by comparing the size of the error bars for each individual subject in Figure 3.3 for the VOs to the conventional letters.

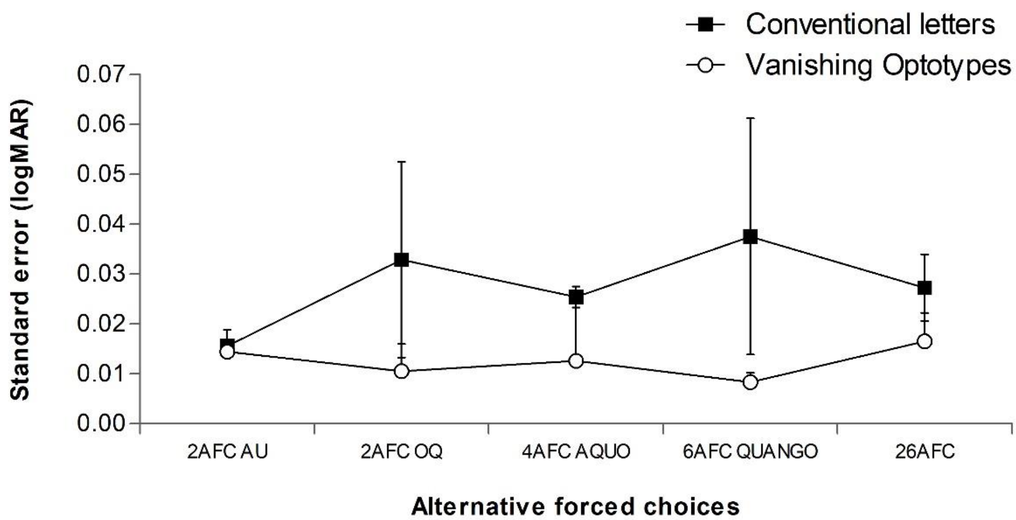


Figure 3.4: Mean standard error of the logMAR thresholds for each nAFC *(adapted from Shah et al., 2011a)*. Values are the average of three repeat readings and across subjects for conventional letters (filled symbols) and VOs (open symbols). Error bars represent the SD (average of subjects).

The graphs in Figure 3.5 display the results for subject NS when repeat measurements were taken using (a) fixed gamma and pThreshold values, (b) different gamma adjusted according to the nAFC with fixed pThreshold values and (c) different gamma and different pThreshold values adjusted to account for the nAFC in the QUEST psychophysical procedure. The range in acuity thresholds obtained for conventional letters and VOs respectively were 0.43 and 0.14 logMAR for (a), 0.37 and 0.11 logMAR for (b) and 0.35 and 0.05 logMAR for (c). A one-way repeated measures ANOVA for each of these conditions revealed significant differences in acuity thresholds between nAFC for the conventional letters with ANOVA results for (a) $F_{4,2} = 49.43$, $p < 0.0001$, (b) $F_{4,2} = 18.72$, $p = 0.0004$ and for (c) $F_{4,2} = 21.70$, $p = 0.0002$. For the VOs, a one-way repeated measures ANOVA revealed significant differences in acuity thresholds between nAFC for (a) and (b) $F_{4,2} = 12.75$, $p = 0.0015$ and $F_{4,2} = 11.47$, $p = 0.0021$ respectively. For (c) ANOVA results revealed no significant differences in threshold $F_{4,2} = 1.64$, $p = 0.2556$. Further examination with a post-hoc Tukey test revealed which acuity thresholds were significantly different (displayed in Table 3.1).

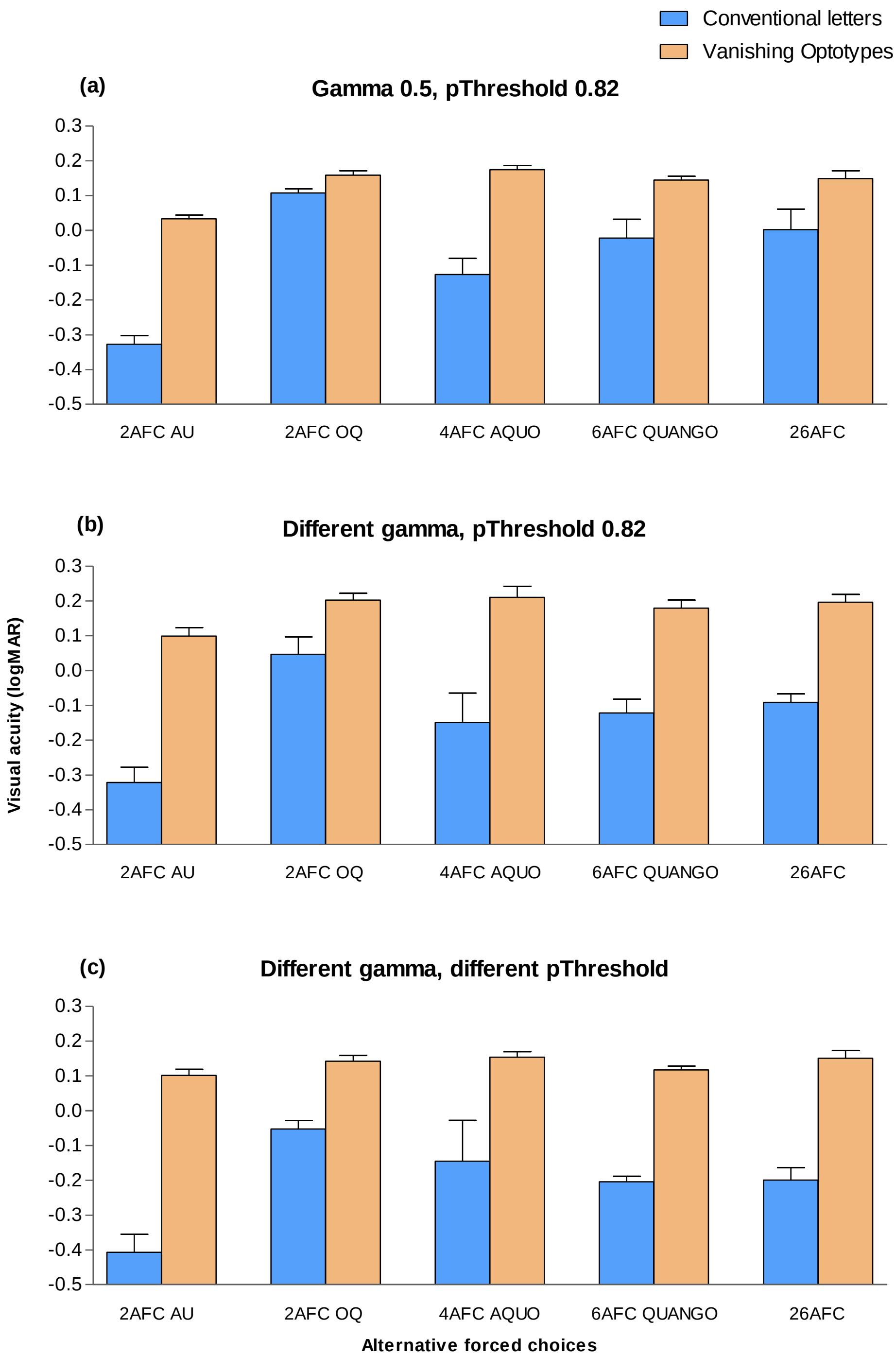


Figure 3.5: Acuity thresholds for conventional letters and VOs with different assumptions in the QUEST psychophysical procedure. Thresholds are the average of 3 readings for subject NS using (a) gamma set at 50% and pThreshold 82%, (b) different gamma adjusted according to the nAFC and pThreshold 82% and (c) different gamma and different pThreshold values adjusted to account for the nAFC. Error bars represent the SD of the three repeat measurements.

	Gamma 50%, pThreshold 82%		Diff. gamma, pThreshold 82%		Diff. gamma, diff. pThreshold	
2AFC AU vs 2AFC OQ	•	•	•	•	•	
2AFC AU vs 4AFC AQUO	•	•	•	•	•	
2AFC AU vs 6AFC QUANGO	•	•	•	•	•	
2AFC AU vs 26AFC	•	•	•	•	•	
2AFC OQ vs 4AFC AQUO	•		•			
2AFC OQ vs 6AFC QUANGO	•		•		•	
2AFC OQ vs 26AFC					•	
4AFC AQUO vs 6AFC QUANGO						
4AFC AQUO vs 26AFC	•					
6AFC QUANGO vs 26AFC						

Table 3.1: Results of a post-hoc Tukey test to investigate under which test conditions acuity thresholds were significantly different. A dot represents the condition for which acuity thresholds were significantly different at $p < 0.05$ for conventional letters (blue) and for VOs (orange).

3.4 Discussion

VA measurements contribute significantly to clinical decision making in assessing disease progression and treatment efficacy. A judgement must be made on any measured change in VA as being clinically significant or not considering the underlying precision and repeatability of the test. Whilst recommendations have been made for ways to reduce TRV, one source of measurement variability investigated in this study is the letter design. The aims in this chapter were to determine the VA measurement repeatability using VO letters and to investigate how acuity thresholds are affected by the number of alternatives available to the subject.

One of the findings of this study is that VA thresholds measured with these high-pass filtered letters are larger than those measured using conventional letters, regardless of the nAFC. This is not unexpected considering the differences in the SF information in the Fourier domain with each letter design. The visual system initially uses the high SF information in conventional letter recognition, until these can no longer be resolved as the letter becomes progressively smaller. Here there is a shift to utilising the low SF information for letter identification in both foveal and peripheral vision (Alexander et al., 1994, Anderson and Thibos, 1999a, Anderson and Thibos, 2004, Bondarko and Danilova, 1997, Chung et al., 2002, Gervais et al., 1984). Since the low SF information is removed in VOs, the visual system is forced to rely on the only available high SF information for letter identification (Majaj et al., 2002) resulting in larger recognition thresholds for these characters. This agrees with the expectations of Watson and Ahumada (2012) as discussed in Section 1.6.1 where their ideal observer model predicted an increase in threshold with letter complexity. The aim of this study was however to compare the effects of different nAFC and threshold variability and not so much to compare absolute differences in threshold letter size between the two target types.

Figure 3.3 demonstrated more similar acuity thresholds with the VOs compared to the conventional letters suggesting that VOs are less affected overall by the nAFC and what these choices actually are. This fits with the original hypothesis that VOs are more equally discriminable than conventional optotypes. As mentioned above and previously also, at threshold, the visual system utilises the low SF information for conventional high contrast letter discrimination. Several studies have indicated large differences in SF content at these low object frequencies (Anderson and Thibos, 1999a, Gervais et al., 1984, Anderson and Thibos, 1999b). Thus as Figure 3.1 illustrated, two letters very different in their SF content such as 'A' and 'U', should remain discriminable down to very small sizes. Conversely, two letters more similar in the low SF content such as 'O' and 'Q', force the visual system to rely on the high SF content for discrimination. Their acuity threshold will thus be larger explaining the large differences found between these two acuity thresholds for conventional letter forms in Figure 3.3. Acuity thresholds for 4, 6 and 26AFC conditions fell in between the two 2AFC conditions with a combination of easy and hard to discriminate letters. However, between-letter differences become smaller and much more uniform if these lower frequencies which give rise to large inter-letter discriminability differences for conventional letters are removed. This is true for the VOs as seen in Figures 3.3 and 3.5 and Table 3.1 where there is no significant difference in performance under the different AFC conditions apart from AU, which even then, has far smaller differences than for the conventional letters. Relying on the higher frequencies only for letter discrimination seems to result in closer between-letter similarity and greater letter uncertainty even under the lower AFC conditions. Perhaps on filtering out the low frequencies, the visual system may even switch to a strategy based more on localised features and less on SF content. It is clear from this study that for conventional letters, the visual system uses different SFs to recognise different letters (Anderson and Thibos, 1999a, Anderson and Thibos, 1999b, Anderson and Thibos, 2004, Bondarko and Danilova, 1997) and these cannot just be regarded as a set of uniformly discriminable stimuli.

Changing parameters of the underlying psychometric function and making certain assumptions about the guess rate had only a slight influence on thresholds achieved (Figure 3.5) and did not affect the overall conclusion, that larger differences between acuity thresholds were found for conventional letters compared to VOs depending on the number of alternatives. Once again, using VOs reduces the effects of differing nAFC on thresholds. Arditi and Cagenello (1993) suggest that acuity scores should be better with a lower compared to higher nAFC (if no correction for guessing is made) since more letters should be read correctly based on chance alone. This is supported by the findings displayed in Figure 3.5 (b) where smaller differences in thresholds are found between different nAFC for the conventional letters after corrections for guessing were made compared to (a) where they were not.

Lastly, measurement variability was found to be lower for the VOs than conventional letters as seen in Figure 3.4. Looking at the error bars also in Figure 3.5, it can be seen that variability was higher for the conventional letters than VOs even under all three testing conditions. Koskin et al., (2007) explain that variability in acuity thresholds attained with conventional letters arises as a result of the existence of two distinctly different thresholds for detection and recognition and the transitional zone between them. Subjects can learn to recognise blurred images close to the detection threshold and this relies on the difference in the low SF content between letters (e.g. 'A' from 'U'). For conventional letters, different low SF content will lead to large inter-letter legibility differences. For any staircase threshold measure, if this difference within steps is significantly greater than between steps, increased threshold measurement variability will result.

In conclusion, the smaller effect of the nAFC on acuity thresholds, combined with advantages of more equal discriminability between letters (resulting in better measurement repeatability), suggests that VOs have potential for use in clinical VA charts. However, there are several questions that need to be answered. The next chapter will examine the relative legibility of each letter in high-pass design to investigate if legibility thresholds do indeed become more similar across the whole alphabet. The behaviour of these letters under non-foveal viewing conditions will also be investigated as will the effect of optical defocus on VO acuity to explore if resolution for these letters is indeed neural sampling limited.

4. Effect of optical defocus on Vanishing Optotype detection and recognition in the fovea and periphery

CHAPTER 4

The work discussed within this chapter has been published in a peer reviewed journal;

Shah N., Dakin S. C. & Anderson R. S. Effect of optical defocus on detection and recognition of Vanishing Optotype letters in the fovea and periphery. Investigative Ophthalmology and Visual Science. 2012;53:7063-70.

This publication was also the feature of a research highlight;

Redmond, T. Toward the optimum measurement of visual acuity. Investigative Ophthalmology and Visual Science. 2012;53:7424.

4.1 Introduction

Conventional letters make excellent refraction targets since their rich SF spectra renders them especially vulnerable to the effects of phase reversals associated with optical defocus (Thorn and Schwartz, 1990, Nestares et al., 2003, Ravikumar et al., 2010). However, large inter-letter differences in their low SF content (Anderson and Thibos, 1999a, Anderson and Thibos, 1999b, Gervais et al., 1984) means that it is often easy to guess the identity of a letter depending, as found in Chapter 3, on the other alternative letter choices available. Much attention has been paid over the years to the letter set and design used in VA charts and how these affect VA scores and repeatability which was discussed in detail in Section 1.4. Whilst the Sloan letter set which is used in the current gold standard VA test, the ETDRS chart, was originally devised to have closely similar letter legibility, closer examination of the data of Sloan et al., (1952), Ferris et al., (1993) and Alexander et al., (1997) indicates that this in fact may not be true. Whilst the letters are arranged such that the average difficulty is the same on each line, Figure 1.5 illustrated the substantial variation in legibility between letters within a line. If the within-line discriminability difference of a test chart is greater than the between-line discriminability difference, the resulting VA thresholds obtained will be very variable with end points spreading over a number of lines. Indeed Carkeet (2001) advises on the introduction of termination rules in the case of a finite nAFC to prevent subjects from attempting subthreshold symbols which could otherwise inflate variances in threshold scores. Whilst the previous chapter demonstrated that VO acuity thresholds were less affected by the nAFC, and furthermore were less variable (Shah et al., 2011a), relative letter recognition thresholds for these optotypes across the whole alphabet has not yet been investigated.

Beyond the fovea, evidence suggests that whilst detection acuity for VOs are limited by the optics of the eye, resolution thresholds may be limited by the RGC sampling density. This comes from the observations of Anderson and Ennis (1999) and Demirel et al., (2012) in which similar thresholds for detection and recognition under normal foveal viewing conditions were found to separate for various pairs of VO targets at 30° eccentricity. Other studies using gratings with the same mean luminance as their background have found similar results (Thibos et al., 1996, Anderson, 2006). Additionally, peripheral grating resolution acuity (but not detection) has been shown to remain robust to the effects of optical blur up to 3 to 4D (Anderson, 1996b, Wang et al., 1997) providing further evidence for the sampling limited nature of resolution for this stimulus. These findings are particularly relevant for certain ocular conditions such as AMD, which result in a loss of foveal function necessitating the uptake of extra-foveal fixation in order to achieve maximum potential VA. However, all these studies examining peripheral grating and VO acuity were conducted only under 2AFC conditions, where the degree of uncertainty is low.

The three goals of this study were firstly to determine relative recognition thresholds of VO letters across the whole alphabet in order to investigate if between-letter performance is indeed less variable than for conventional letters. Secondly, to examine how detection and recognition thresholds for individual letters vary under foveal and peripheral viewing and conditions of higher uncertainty (26AFC). This should allow us to determine which individual letters display sampling limited performance. The final goal of this study was to explore how these thresholds vary under differing levels of defocus since the relative vulnerability of VO letters to the effects of optical defocus remains unknown.

4.2 Methods

4.2.1 Subjects

Two healthy experienced psychophysical observers (NS aged 27 years and RSA aged 45 years) with no ocular abnormalities were recruited for this study. Both subjects had VA of 6/5 or better and the refractive error was corrected for foveal and extra-foveal testing at 10° eccentricity in the nasal field of the right eye using the methods described in Section 2.1.5. This eccentricity was chosen for a number of reasons. Patients with AMD often adopt a preferred retinal locus (PRL) (Crossland et al., 2005a, Crossland et al., 2011b) in which an eccentric retinal area away from a central scotoma is used for fixation. Although this position has been shown to vary with the task in question (Crossland et al., 2011a) it is not uncommon for a PRL to be located around this eccentricity (Rubin and Feely, 2009) and thus useful clinical information can be gained by testing at this location. Furthermore, at this position larger defocused letters did not encroach on the fovea but, at the same time, the peripheral refractive error differed little from that of the fovea at this location (subject NS -0.25DS fovea, 0DS periphery and subject RSA +0.50DS for both fovea and periphery).

A naïve observer (SS) aged 24 years acted as a control when testing under zero-blur conditions. Refractive error in the test eye for this subject was +0.50DS/-2.50DC x 15 under foveal conditions and +0.75DS/-2.50DC x 15 under peripheral conditions. Ethical approval was obtained for this study from the relevant UCL Research Ethics Committee and all procedures adhered to the tenets of the Declaration of Helsinki. All test examinations were conducted in the Spatial Vision Laboratory at the UCL Institute of Ophthalmology, London, UK.

4.2.2 Procedure

VO letters (described in Section 2.1.2) and conventional letters of a 5 x 5 unit grid design were generated using MATLAB (version 7.6, Mathworks, Inc., Natick, MA, USA) and were presented using an Apple Macintosh computer (Apple, Inc., Cupertino, CA) on a γ-corrected high resolution (1280 x 1024 pixels) Dell Trinitron P992 CRT monitor (Dell Corp. Ltd, Bracknell, Berkshire, UK). A Bits++ video processor (Cambridge Research Systems, Ltd., Rochester, UK) was used to achieve true 14-bit contrast resolution whilst the OpenGL capabilities of the computer's built-in graphics card (ATI Radeon X1600; AMD, Sunnyvale, CA, USA) was used to scale the stimuli. This bilinear interpolation procedure allowed stimuli to be displayed of arbitrary size with sub-pixel resolution whilst retaining their balanced luminance structure. The luminance of the background of the CRT monitor measured 53 cd/m^2 for the VO letters. The white background for the conventional black-on-white (B/W) letters had a luminance of 113 cd/m^2 to yield the same on-screen contrast as the VO letters of 98%. Control measurements were also taken using conventional design white letters on the same grey background (W/G) as the VO letters for the zero defocus condition in order to investigate any potential influences on acuity estimates that a higher background luminance may have. Thus, whilst this achieved the same retinal illuminance as for VO letters, it resulted in a lower on-screen contrast of 35%. All testing was conducted under low room illumination to avoid screen reflections. A viewing distance of 8 m was used for all foveal testing in order that the VO letters could vanish without pixilation effects for detection thresholds, whilst all peripheral testing was conducted at 1.6 m in order to ensure that the letter could fit on the screen under higher levels of optical defocus. The screen subtended 11.6° x 9.8° and one pixel subtended 0.55 arcmins at this near distance.

Threshold VA was determined for each of the 26 letters of the alphabet for each subject for both conventional and high-pass letters with baseline refraction fully corrected using full aperture trial lenses. For the two experienced and one naïve observer, detection and recognition thresholds were determined for the two conventional optotypes (B/W and W/G) and for the VO letters under foveal and peripheral viewing conditions. Detection and recognition tasks were conducted in separate runs, thus each subject was tested under 12 conditions in total.
In the second stage of the study, recognition thresholds only for the B/W letters and detection and recognition thresholds for the VO letters were determined for the two experienced observers under different levels of defocus, increasing from their initial refraction in +1D steps. The maximum level of defocus under foveal viewing conditions was +3D, beyond which the letters became too large to be generated on the screen at the test distance employed, and a maximum of +7D under peripheral viewing conditions.

An ascending MOL was employed to determine thresholds for each test condition. For the detection task, the observer moved the computer mouse to progressively increase the size of a letter displayed initially at subthreshold size, indicating the point at which any contrast was just visible against the background by clicking the mouse button. In the recognition task, the same procedure was applied but the observer was this time required to verbally report the letter identity to the examiner when just discernible, before clicking on the mouse to register this letter size as the recognition threshold if the answer was correct. In the case that the identity was incorrectly reported, the observer continued to increase the size until able to correctly identify the letter. This procedure was repeated five times for the detection and recognition task for each letter of the alphabet which were presented in a randomly interleaved order with viewing time unrestricted.

There were several reasons to employ an ascending MOL rather than forced choice procedures. An individual-letter comparison experiment would require 26 different interleaved staircases which would result in more than 800 presentations per run. Observer fatigue would introduce variability, particularly if this was undertaken for detection (two intervals per presentation) as well as recognition for each of the blur conditions both in the fovea and the periphery. The previous study also demonstrated that circular letters, such as 'O' and 'Q' are commonly confused with each other, behaving as a separate subset within a 26AFC test. Thus, if an observer is able to determine that a letter belongs to this group of round letters but is unable to certainly identify which one it is, under forced choice conditions (effectively a 4AFC subset), a bias towards a particular letter (e.g. 'O') often occurs. This will falsely boost the threshold for that particular letter. A truer threshold is achieved under MOL procedures since the letter must be increased in size until the observer can more confidently identify the letter rather than forcing a decision. Furthermore, there were no prior expectations as to which, if any, of the 26 letters may display differences in their detection and recognition thresholds to influence the experienced observer's criterion due to the lack of previous similar studies on high-pass filtered letters. Nevertheless, an adaptive forced-choice reversal staircase procedure (QUEST) in the zero-defocus condition for one of the trained observers (NS) was employed for control and comparison purposes in order to assess the possibility that a criterion-based MOL procedure may permit bias of the results in non-naïve observers. In the QUEST procedure, the prior density function was limited by the minimum and maximum displayable letter size on the screen. The slope (β) of the psychometric function used was set to 3.5 with gamma and pThreshold set to 50% and 75% respectively for the detection task and 3.8% and 52% respectively for the recognition task.

4.2.3 Statistical analysis

The final threshold letter size, generated as a percentage by reference to a box size of 512 pixels, was converted to a logMAR score for further analysis where the stroke width for the VO included both the central dark bar and the surrounding white flanks as described in Section 2.1.2. The GraphPad Prism statistical analysis package (GraphPad Software, Inc., La Jolla, CA) was used to compare VA thresholds using the one-way repeated measures ANOVA and statistically significant results were investigated using Bonferroni's Multiple Comparison post-hoc analysis for selected pairwise comparisons.

4.3 Results

The detection and recognition thresholds for the high-pass letters and recognition values for B/W conventional letters under increasing levels of optical defocus, up to +3D in the fovea and up to +7D at 10°, are displayed in Figure 4.1 (a) and Figure 4.1 (b) respectively. Each point is the average across the 26 letters and both subjects with error bars representing the SD.

Differences in the three thresholds obtained in the fovea (Figure 4.1a) under different levels of blur (up to +3D) were tested for significance using a one-way repeated measures ANOVA, $F_{11,25} = 1382$, $p < 0.0001$. A post-hoc Bonferroni test confirmed significantly 'better' recognition acuity thresholds for conventional B/W letters than either detection or recognition thresholds for the VOs at zero blur at $p < 0.05$ (-0.01 logMAR vs. 0.11 and 0.14 logMAR respectively). The difference in detection and recognition values for VOs under foveal conditions with 0D blur did not reach statistical significance and these thresholds both increased steadily with optical defocus, separating with increasing blur. Whilst recognition thresholds for VOs changed on average by 0.28 logMAR/D, recognition thresholds for conventional letters changed more rapidly by 0.35 logMAR/D. Thus with +1D of blur, conventional and VO recognition thresholds were similar, beyond which recognition thresholds for conventional letters were worse than for VOs at $p < 0.05$.

A one-way repeated measures ANOVA, $F_{23,25} = 777.6$, $p < 0.0001$ and a post-hoc Bonferroni test was again used to investigate differences in thresholds with up to +7D optical defocus at 10° in the periphery (Figure 4.1b). Recognition thresholds for conventional letters were statistically significantly lower than for VOs with 0D at $p < 0.05$. These two thresholds increase approximately in parallel, remaining significantly different until +7D at which point they converge somewhat but still remain statistically different. Unlike in the fovea, recognition thresholds were significantly higher at $p < 0.05$ than detection thresholds for the VOs at 10° in the periphery under 0D blur (0.85 vs. 0.56 logMAR). Both thresholds increase steadily with blur, with recognition thresholds less affected than detection thresholds (changing by 0.09 logMAR/D vs. 0.12 logMAR/D respectively) over the full +7D blur range.

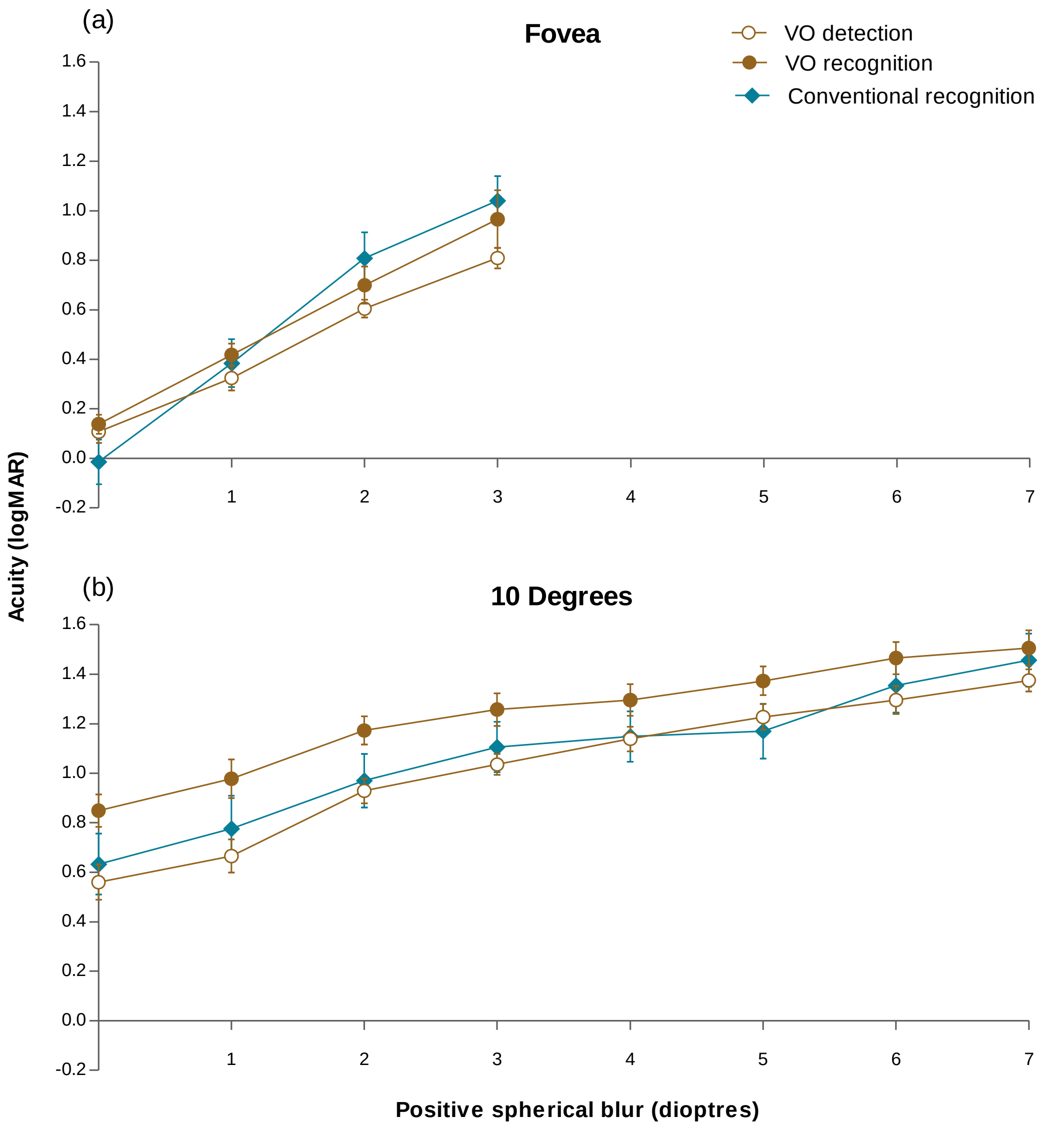


Figure 4.1: Detection and recognition thresholds for the VOs, and recognition values for the conventional B/W letters under different levels of optical defocus in the fovea and at 10° eccentricity *(adapted from Shah et al., 2012a)*. Thresholds are the average across the 26 letters and both experienced observers in (a) the fovea and (b) 10°. Error bars represent the SD of the 26 letter thresholds (mean of both subjects).

Figure 4.2 displays the detection and recognition values under 0D optical defocus for the individual letters of (a and b) B/W conventional letters, (c and d) VOs and (e and f) W/G conventional letters in the fovea (top row) and periphery (bottom row). The upper half of Figure 4.2 (a and b) displays recognition thresholds for the individual conventional B/W letters under the maximum +3D and +7D blur conditions in the fovea and periphery respectively. The upper half of Figure 4.2 (c and d) shows both detection and recognition thresholds for the individual VO's under +3D and +7D blur conditions in the fovea and periphery respectively. Thresholds are calculated as the average of both experienced observers, with error bars representing the standard error of the five threshold measurements for each letter (average of both subjects).

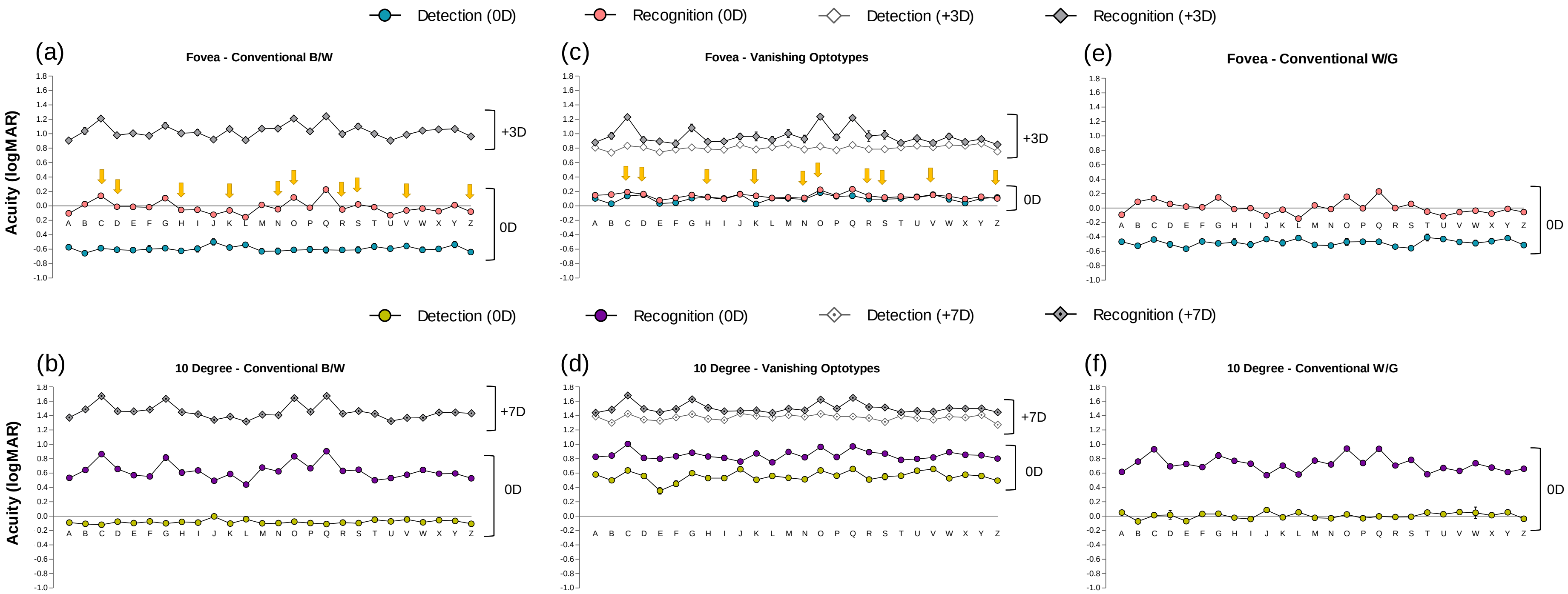


Figure 4.2: Detection and recognition values for each letter of the alphabet in the fovea and periphery for the conventional and VO letters ***(adapted from Shah et al., 2012)*****.** Thresholds are the average of both experienced observers for the (a, b) B/W conventional letters, (c, d) VOs and (e, f) W/G conventional letters in the fovea (top) and periphery (bottom). Thresholds under the maximum blur conditions (+3D fovea, +7D periphery) where measured are also shown. Error bars represent the standard error of the five measurements for each letter (average of both subjects). The arrows indicate recognition thresholds for the set of 10 Sloan letters.

The large difference between the mean luminance of the letters and the background for the conventional B/W letters resulted in, as expected, lower detection than recognition thresholds for both subjects in the fovea under 0D of optical defocus (mean -0.59 logMAR versus -0.01 logMAR). Significant between-letter recognition threshold differences are evident, the highest thresholds observed for the circular letters (C G O Q). The variation in thresholds across all 26 letters was higher for recognition than for detection (SD 0.09 vs. 0.04 logMAR). The only observable difference for the conventional W/G letters (Figure 4.2e) compared to the higher contrast B/W letters under 0D defocus was slightly raised letter detection thresholds (mean -0.47 rather than -0.59 logMAR). Performance for the conventional W/G letters was otherwise qualitatively and quantitatively very similar to the B/W letters with mean recognition thresholds of 0.01 logMAR and SD of 0.09 and 0.05 logMAR for recognition and detection thresholds respectively.

Whilst it is known that under foveal viewing conditions VOs display closely similar average detection and recognition performance, Figure 4.2 (c) indicates that this is also applicable on an individual letter basis. Consequently, the SD of the between-letter differences for VOs is similar for both recognition and detection (0.04 and 0.05 logMAR respectively), which is comparable to the SD found for conventional letter detection but far lower than that for conventional letter recognition.

Under peripheral viewing at 10° eccentricity (Figure 4.2b), differences between detection (mean -0.08 logMAR) and recognition (mean 0.63 logMAR) were observed to be even larger for the conventional B/W letters compared to the fovea, with the circular letters again almost behaving as a separate subset displaying the highest recognition thresholds. Slightly greater variation was found in between-letter recognition thresholds (SD 0.12 logMAR) compared to the fovea whilst similar variation was found for detection (SD 0.03 logMAR). The conventional W/G letters again displayed very similar qualitative performance to conventional B/W letters with slightly higher detection and recognition thresholds of 0.02 logMAR (SD 0.06 logMAR) and 0.73 logMAR (SD 0.11 logMAR). The VO's demonstrate considerable differences between detection and recognition thresholds for all letters in the periphery (Figure 4.2d) unlike in the fovea, the magnitude of this difference dependent on the letter under consideration. Whilst detection and recognition threshold variation between letters was higher than in the fovea for the VOs (SD 0.07 logMAR for both detection and recognition), the recognition threshold variation was still markedly lower than for conventional letters at this same location. The detection and recognition thresholds and SDs (average of all 26 letters and both experienced observers) for each letter design for the fovea and periphery under 0D optical defocus are summarised in Table 4.1.

Acuity logMAR	B/W Detection	B/W Recognition	VO Detection	VO Recognition	W/G Detection	W/G Recognition
Fovea	-0.59 ± 0.04	-0.01 ± 0.09	0.11 ± 0.05	0.14 ± 0.04	-0.47 ± 0.05	0.01 ± 0.09
Periphery	-0.08 ± 0.03	0.63 ± 0.12	0.56 ± 0.07	0.85 ± 0.07	0.02 ± 0.06	0.73 ± 0.11

Table 4.1: Summary of the detection and recognition thresholds and SDs for all letter types under 0D blur conditions in the fovea and periphery *(adapted from Shah et al., 2012a).* Data are the average of all 26 letters and both experienced observers.

Looking at the results under dioptric blur (Figure 4.2a and b top part), recognition performance decreased on average by 1.05 logMAR under +3D blur in the fovea and by 0.83 logMAR with +7D blur in the periphery for conventional B/W letters. However, the between-letter variability was very similar to 0D blur (SD 0.10 logMAR with +3D in the fovea and 0.11 logMAR with +7D in the periphery). The circular letters continue to behave as a separate subset in both cases.

Interestingly, the detection and recognition thresholds actually separate with blur for the VOs in the fovea (Figure 4.2c top part) most noticeably again for the circular letters. Detection performance decreased on average by 0.70 logMAR with +3D blur in the fovea compared to 0.83 logMAR for recognition thresholds in this range. In the periphery (Figure 4.2d top part), whilst detection and recognition performance both declined with blur, detection performance was most affected (0.82 logMAR deterioration with +7D blur compared to 0.66 logMAR for recognition over this range). Consequently, the 0.29 logMAR average difference in detection and recognition thresholds at 0D blur in the periphery reduced to 0.13 logMAR at +7D blur.

Figure 4.3 displays the detection and recognition values in the fovea and recognition values in the periphery for the conventional B/W letters and detection and recognition values for the VOs in the fovea and periphery under 0D blur for the naïve subject using an ascending MOL. It can be seen that qualitatively, the results are similar to those described so far for the experienced observers.

Figure 4.4 shows the detection and recognition values for both the conventional B/W letters and VOs in the fovea and periphery under 0D blur for the experienced observer (NS) using a forced-choice QUEST psychophysical staircase procedure. A larger difference was found between mean (average for the 26 letters) detection and recognition thresholds for the VO letters in the fovea using the staircase compared to ascending MOL procedure for subject NS (0.10 logMAR vs. 0.00 logMAR). Interestingly, recognition thresholds attained for the conventional letters differed more between the two methodologies than those for the VOs for subject NS. Recognition thresholds for the conventional letters were 0.18 and 0.13 logMAR smaller for the fovea and periphery respectively compared to 0.01 and 0.02 logMAR smaller for the VOs with the staircase procedure. This may be attributable to the fact that that there is greater uncertainty with the VOs which must thus be of a particular size before they can be correctly identified even using a forced choice procedure.

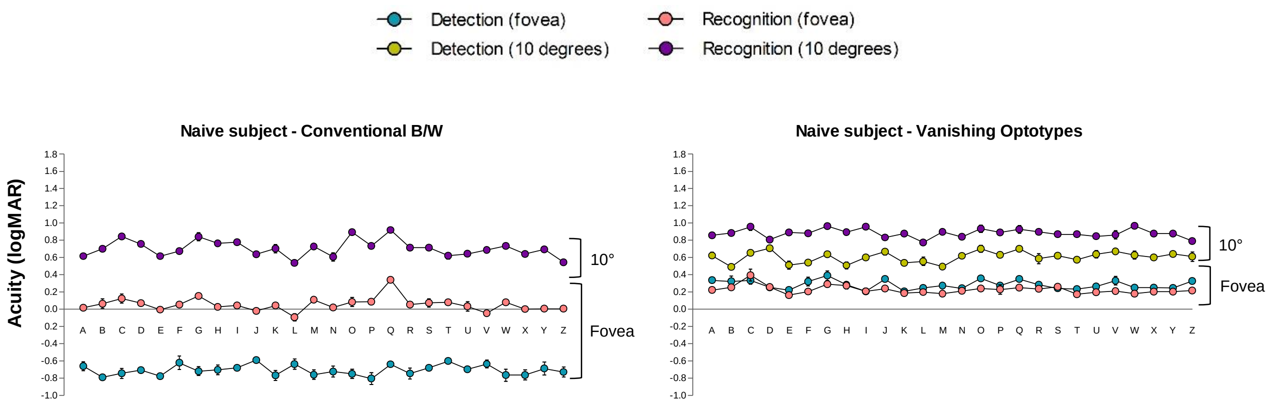


Figure 4.3: Foveal and peripheral thresholds for the naïve subject using an ascending method of limits with 0D optical defocus. The graph on the left displays detection and recognition values for the conventional B/W letters in the fovea and recognition values in the periphery. The graph on the right displays detection and recognition values for the VOs in the fovea and periphery. Error bars represent the standard error of the five threshold measurements for each letter.

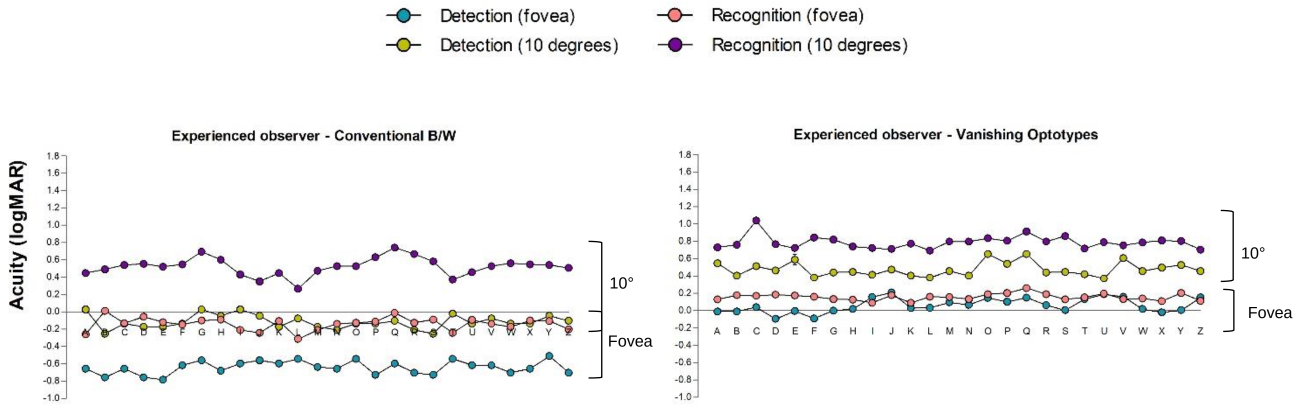


Figure 4.4: Foveal and peripheral thresholds for the experienced observer using a forced-choice QUEST psychophysical staircase procedure with 0D optical defocus. Detection and recognition values for the conventional B/W letters are shown on the left and for the VOs on the right in the fovea and periphery. Error bars represent the standard deviation of the threshold measurements.

4.4 Discussion

The previous study (Chapter 3) demonstrated that VOs elicit less variable acuity thresholds that were also less affected by the nAFC available than conventional letters (Shah et al., 2011a). The reason for this is thought to lie in the absence of the low SF components which vary greatly between conventional letters, rendering the VO letters more equally resolvable. This study aimed to look at the separate detection and recognition thresholds for individual letters of VO design in order to determine if between-letter performance is indeed less variable than for conventional letters. This was done for both foveal and extra-foveal viewing to determine differences between detection and recognition thresholds at these two locations in order to investigate which letters in particular, display sampling limited performance and to look at these thresholds under different levels of optical defocus.

Looking closely at the individual letters in Figure 4.2 (a), considerable differences were found in the recognition thresholds for the conventional letters under zero defocus, the inter-letter range being 0.40 logMAR. The poorest acuity thresholds were found for the circular letters (C G O Q). The similarity of these letters to one another and strong dissimilarity to the rest of the letter set mean that they appear to behave as a separate subset. Indeed, Banister (1927) discovered that the mistakes the subject makes in identifying a letter often group themselves into letters which have the same general outline and found one of the largest classes to be C D G O Q which required a higher level of acuity to distinguish between than letters belonging to different groups. The vertical arrows on Figure 4.2 (a) identify the 10 Sloan letters and if only these letters are considered, the legibility range reduces to 0.22 logMAR (SD 0.08 logMAR, mean of the two experienced subjects). This is similar to the legibility differences found in other studies (Alexander et al., 1997) but is far larger than the 0.10 logMAR step size between lines on a conventional logMAR chart. Section 1.4.2 discussed in detail the possible contribution that the within-line differences in Sloan letter legibilities may have on reported measurements of test variability and if these are larger than the between-line differences, acuity measurement variability results. McMonnies and Ho (2000) discovered that chance combinations of easy or difficult letters can lead to significant variations in line-difficulty. This limits the ability of acuity tests to support reliable detection of small losses or gains in visual function (with disease progression or improvement with treatment). The VOs (Figure 4.2c) elicited closely similar detection and recognition thresholds for every letter and displayed a much lower inter-letter variation in recognition thresholds than the conventional letters under zero defocus in the fovea, with a range of 0.15 logMAR across all 26 letters. In addition, the SD of VA measurements was much lower with the VO compared to conventional letters (0.04 vs. 0.09 logMAR). Considering the Sloan letter set once again, the between-letter variation reduced slightly to 0.13 logMAR (SD 0.06 logMAR, mean of the two experienced observers) although this is a very small difference.

The previous study (Shah et al., 2011a) demonstrated lower variability in acuity measurements with VO compared to conventional letters and the results of this present study suggest that this is likely due to the smaller inter-letter legibility differences. It may thus be desirable to select an alternative letter set to the Sloan set to use on conventional letter charts which have more closely similar legibility and Grimm et al., (1994) have summarised some studies which have attempted to do this. Furthermore, VO letters could be more appropriate targets with which to construct acuity charts since letter choice should have less effect on VA thresholds than found with conventional letters, thus reducing measurement variability. The circular VO letters (C G O Q) did not behave so much as a separate subset as they did for conventional letters under zero defocus in the fovea but did so as defocus increased (upper part of Figure 4.2c). The explanation for this may be that the low-pass filtering effects of blur lenses attenuate the higher frequencies which is where the information required to distinguish such similar letters lies. Thus the inclusion of these items on a VO letter clinical test chart could potentially increase acuity measurement variability under higher levels of defocus as they become more difficult to tell apart.

Under peripheral viewing at 10°, the inter-letter recognition differences were once again large for the conventional letters (0.46 logMAR) with the circular letters still behaving as a separate subset (Figure 4.2b). From Figure 4.2 (d), it can be seen that these differences were again much smaller for the VOs (0.26 logMAR). Unlike in the fovea however, significant differences between detection and recognition thresholds for all of the individual VO letters was found, indicating that these letters do not 'vanish' extra-foveally. As discussed in Section 1.6, for targets with the same mean luminance as their background, a superiority of detection acuity over recognition or resolution acuity is strong evidence that the resolution task is limited by retinal sampling rather than the eye's optics. This agrees with the findings of previous studies using high-pass targets but with a lower number of alternatives (Anderson and Ennis, 1999, Demirel et al., 2012). The VO letters showing the least difference between these thresholds (0.25 logMAR or less) were the letters A, D, J, L, T, U and V.

In agreement with the previous study (Shah et al., 2011a), the foveal acuity thresholds were larger for the VOs than the conventional letters under zero defocus, observable in Figure 4.1 and Figure 4.2. As explained previously in Chapter 3, removing the low SF information means that letter size must be increased so that the higher object frequencies become lower in retinal frequency in order for the visual system to be able to use these frequencies for recognition. However, somewhat surprisingly, the threshold letter size for conventional letter recognition was slightly, but significantly larger than for VOs beyond +1D of defocus. With progressive defocus, it is the higher retinal frequencies that first start to phase-reverse. Some studies have suggested that this results in the masking of the lower frequencies thus making the letters increasingly harder to resolve (Thorn and Schwartz, 1990). Other computational studies however have found that this may not necessarily be the case. Akutsu et al., (2000) reported little effect on defocused letter VA on removing the spectrum above the first cut-off of the OTF. Ravikumar et al., (2010) found that, in the presence of positive spherical aberration, the impact on VA with positive defocus was not significantly different for standard, phase-rectified or low-pass filtered defocus. This led them to conclude that the primary cause of acuity loss with conventional letters for positive blur was contrast reduction. For VO letters, a possibility is that phase reversal of the higher frequencies composing the lighter edges of the stimuli, causes the edges to become darker, effectively making the letter strokes thicker and the letter thus more discriminable.

Figure 4.1 (b) demonstrates again a difference in detection and recognition thresholds for the VOs in the periphery and these letters behave in a similar manner to that found in previous studies which employed peripheral gratings with the same mean luminance as their surround. The effect of optical defocus was found to be substantially less for both letter types in the periphery than the fovea, with a logMAR/D loss of only approximately a quarter of that seen in the fovea. Recognition acuity with VOs in this study was not quite as robust to blur as seen in previous studies using gratings (Anderson, 1996b, Wang et al., 1997). This may be because retinal sampling may not be the only significant limiting factor involved in peripheral viewing, unlike for gratings. Under optical phase reversal, whilst the appearance of gratings remains largely veridical, the spurious appearance of the letters with increasing defocus renders them more ambiguous to the underlying RGC mosaic. Additionally, there remains a difference between the detection and recognition thresholds for high-pass letters all the way up to +7D, whilst previous studies have found that grating detection and resolution performance are identical by around 3 to 4D defocus. Again, this is likely because the spuriously defocused letter allows contrast detection that may not be adequately veridical to accurately resolve. This also seems to occur in the fovea Figure 4.1 (a), where thresholds separate somewhat for VOs as defocus increases.

The separation of detection and recognition thresholds for VOs in the periphery and their relative robustness to optical defocus suggests a sampling limited performance for recognition but not detection of these letters in peripheral vision. In HRP, the measured detection threshold of a ring VO is assumed to also represent the resolution threshold for this target to give a sampling limited performance measure. This present study however suggests that this assumption may be inaccurate since results for the letter 'O' in Figure 4.2 (d) demonstrate significantly different detection and recognition thresholds in the periphery. In conditions such as AMD which result in a loss of foveal photoreceptors, recognition thresholds with high-pass letters may become sampling limited in a similar fashion to that found in the periphery and a separation of detection and recognition thresholds for these optotypes may be a useful early sign of the condition. Age-related optical factors behave differently to optical defocus and the next chapter will investigate the effect of these on VO recognition thresholds before further clinical work to determine the role of high-pass letters in AMD is piloted.

5. Effect of simulated lens opacity on recognition thresholds for Vanishing Optotypes and conventional letters

CHAPTER 5

5.1 Introduction

When considering the development of test/s using VOs, which may be more specific and sensitive to neural damage along the visual pathway, it is essential to determine their robustness to aberrations and imperfections in the optics of the eye. This is important to allow accurate discrimination between the functional consequences of optical and neural losses. The previous chapter examined the performance of VOs with optical defocus. However, the processes behind image degradation differ for age-related lens ageing and refractive error, namely light scatter and absorption versus optical defocus. Lens opacities or cataracts cause a range of symptoms (Lee et al., 2005) including blurred vision, monocular diplopia, refractive changes and colour vision impairment. These are caused by changes in forward light scatter (Van Den Berg et al., 2007), light absorption and spectral transmission characteristics (Artigas et al., 2012) and an increase in higher order aberrations (Zhu et al., 2016, Rocha et al., 2007). One of the commonest reported symptoms however is glare, particularly as a consequence of viewing car headlights (Babizhayev, 2003). This is due to an increase in retinal stray-light with natural ageing of the crystalline lens, increasing as cataract develops (Van Den Berg et al., 2007, de Wit et al., 2006, van der Meulen et al., 2012). This retinal stray-light, caused by forward light scattering, produces a veiling luminance on the retina resulting in the reduction of retinal image contrast (Elliott et al., 1991, Elliott et al., 1996) and, if high enough, glare.

A 'cataract causing visual impairment' was historically defined as a VA of worse than 6/12 in one or both eyes where the impairment was attributable to lens opacity (Minassian et al., 2000). However, it is now widely acknowledged that on its own, VA measured with high-contrast black-on-white letter targets can often underestimate the visual disability caused by cataract (Amesbury et al., 2009, Van Den Berg et al., 2007, Elliott et al., 1990, Zhu et al., 2016, Elliott et al., 1996). Indeed Quality of Life Questionnaires - designed to measure vision-related functional status - provide information on functional impairment from cataract which VA measures do not convey (Steinberg et al., 1994, Amesbury et al., 2009, Javed et al., 2015). A number of studies have demonstrated that assessment of other components of visual function, such as spatial contrast sensitivity (CS) (Bal et al., 2011, Superstein et al., 1997, Elliott et al., 1996) and disability glare (Rubin et al., 1993, Superstein et al., 1997, Elliott et al., 1996), provides important information about visual quality, accounting for complaints of poor vision in those where VA still remains normal. Hess and Woo (1978) investigated contrast thresholds for a range of different SFs in subjects with monocular cataract and found that those in whom CS is reduced across all frequencies, (typically associated with late grade cataract (Chua et al., 2004)), is attributed to wide angle scatter and these are the patients who complain of significant visual problems despite maintaining a good VA (de Waard et al., 1992). Indeed, it has been reported that measurements of binocular CS correlate more strongly with a patient's perceived visual disability than measurements of binocular VA (Elliott et al., 1990). Cheng et al., (2013) demonstrated a stronger correlation between CS and cataract severity, graded using the Lens Opacities Classification System III (LOCS III), than between logMAR VA and LOCS III. In light of this, and coupled with advancements in techniques offering improved risk-to-benefit ratios, the Commissioning Guidance from the Royal College of Ophthalmologists now recommends consideration be given to cataract surgery when disabling visual symptoms attributable to cataract interfere with daily living activities (The Royal College of Ophthalmologists, 2015).

Objective measures of cataract include slit lamp examination, (which is dependent on back scatter (Elliott and Bullimore, 1993), and the widely used LOCS III for grading cataracts (which is somewhat subjective). More recent studies have suggested that measurements of stray light may offer better objective measures with which to quantify cataract and form part of preoperative clinical considerations (van der Meulen et al., 2012, Van Den Berg et al., 2007, Bal et al., 2011, Galliot et al., 2016). Elliott et al. (1991) found better correlation of CS than VA scores with light scatter at different angular distances from the patient's eye, measured with a stray light meter. A computerised stray light meter, the C-Quant (Oculus) objectively measures the amount of intraocular forward scattered light giving valid, highly repeatable results (Van Den Berg et al., 2007, Cervino et al., 2008).

In considering VOs in the development of acuity tests which are more specific and sensitive to neural damage, it is important to investigate the impact of optical degradations caused by lens ageing, on VO acuity thresholds. The present study aimed to investigate the relative effect of using white opacity-containing filters to induce different levels of wide-angle light scatter (Zlatkova et al., 2006) to simulate lens ageing, measured using the C-Quant, on recognition thresholds for VOs and conventional letters in the fovea and the periphery which is unknown. The methodology has been validated by Anderson et al., (2009) and Bergin et al., (2011) in investigating the robustness of various perimetry tests to different levels of induced intraocular stray light using these filters.

5.2 Methods

5.2.1 Subjects

One healthy experienced psychophysical observer (NS aged 31 years) and one naïve subject (JM aged 29 years) with no ocular abnormalities were recruited for this study. Both subjects had VA of 6/5 or better with refractive error corrected for foveal (subject NS -0.25DS and subject JM -3.75DS/-0.50DC x 180) and extra-foveal testing at 10° eccentricity in the nasal field of the right eye (subject NS 0DS and subject JM -3.50DS/-0.50DC x 180) using the methods described in Section 2.1.5. Ethical approval was obtained for this study from the relevant UCL Research Ethics Committee and all procedures adhered to the tenets of the Declaration of Helsinki. All test examinations were conducted in the Spatial Vision Laboratory at Moorfields Eye Hospital, London, UK.

5.2.2 Procedure

VO letters (described in Section 2.1.2) and conventional letters of a 5 x 5 unit grid design were generated using MATLAB (version 7.6, Mathworks, Inc., Natick, MA, USA) and were presented using an Apple Macintosh computer (Apple, Inc., Cupertino, CA) on a γ-corrected high resolution (1280 x 1024 pixels) Dell Trinitron P992 CRT monitor (Dell Corp. Ltd, Brackness, Berkshire, UK). True14-bit contrast resolution was achieved using a Bits++ video processor (Cambridge Research Systems, Ltd., Rochester, UK) whilst scaling of the stimuli was achieved using the OpenGL capabilities of the computer's built-in graphics card (ATI Radeon X1600; AMD, Sunnyvale, CA, USA). The background luminance of the CRT monitor for the VOs measured 53.3 cd/m^2 whilst the white background for the conventional black-on-white letters had a luminance of 113.2 cd/m^2 thus yielding the same on-screen contrast as the VO letters of 98%. Stimuli were presented for 500 ms. All testing was conducted under low room illumination to avoid screen reflections with a viewing distance of 8 m for all foveal testing and 1.6 m for all peripheral testing. The screen subtended 11.6° x 9.8° and one pixel subtended 0.55 arcmins at this near distance.

Recognition threshold VA was determined for each subject, for both conventional and VOs, in the fovea and periphery, using a 10 AFC (Sloan letter set). These measurements were repeated three times for 6 different levels of induced stray light in a random order using an adaptive forced-choice reversal staircase procedure (QUEST). The initial letter size displayed was 115.8 x 115.8 arcmins and the prior density function limited by the minimum and maximum displayable letter size on the screen. The slope (β) of the psychometric function used was set to 3.5 with gamma and pThreshold set to 10% and 75% respectively. Each test run involved 50 letter presentations in total with the final acuity threshold determined by QUEST's built in maximum likelihood estimation procedure of threshold. Subjects were made aware of the letter set available and the subject's verbal response as to the letter identity was entered by the examiner on the keyboard.

5.2.3 Filters

Five (Filters 1 to 5) white resin opacity-containing filters (LEE Fog Filters, Andover, UK) were used to simulate the increasing light scatter and absorption with age. Previous work by Zlatkova et al., (2006) has confirmed that wide-angle scatter is being generated by these filters and they display very flat spectral transmission spectra, thus representing a good simulation of at least some kinds of cataract (Anderson et al., 2009). In this work, they offer a means of assessing how VO acuity is affected by image degradation caused by scatter as opposed to refractive blur.

A computerised stray light meter (C-Quant; Oculus GmbH, Wetzlar, Germany) was used to measure the baseline intraocular stray light (no fog filter), using the psychophysical compensation comparison method (Franssen et al., 2006) for both subjects and the individual increase in forward intraocular stray light when each of the filters (Filters 1 to 5) were placed in front of the right eye close to the cornea. Values are expressed as log [stray light parameter] (log[s]) with higher values indicating greater levels of stray light.

Figure 5.1 demonstrates how the stray light values for the experienced (NS) and naïve subject (JM) changed for each filter. Baseline measures with no filter were within the normal expected range given the age of both subjects. An assessment of the reliability of the test is provided by the C-Quant stray light meter and all measurements for both subjects were found to be within acceptable reliability parameters with expected SD ≤0.08 log units and reliability coefficient (Q) ≥1. It can be seen that the white resin opacity-containing filters progressively increase the forward light scatter for each subject as expected.

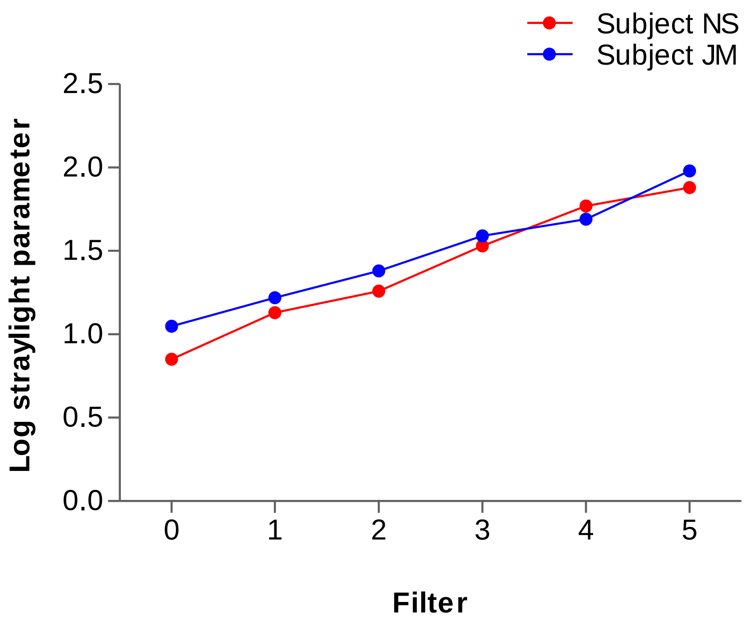


Figure 5.1: Stray light values for each filter measured using the C-Quant stray light meter. Baseline stray light value (Filter 0) and stray light values for each filter (Filters 1 to 5) are shown for the experienced (NS) and naïve (JM) subjects.

Figure 5.2 demonstrates how average stray light values increase with age as recorded by the C-Quant stray light meter. The stray light values with each filter are also plotted for the experienced observer (NS aged 31 years) to demonstrate how these increase from baseline. Filters 1, 2 and 3 increase the stray light levels of an average 31 year old to the stray light levels of a 62, 72 and 90 year old respectively. The last two filters (4 and 5) take the stray light value outside the scale and demonstrate stray light levels expected with significant cataract (de Wit et al., 2006).

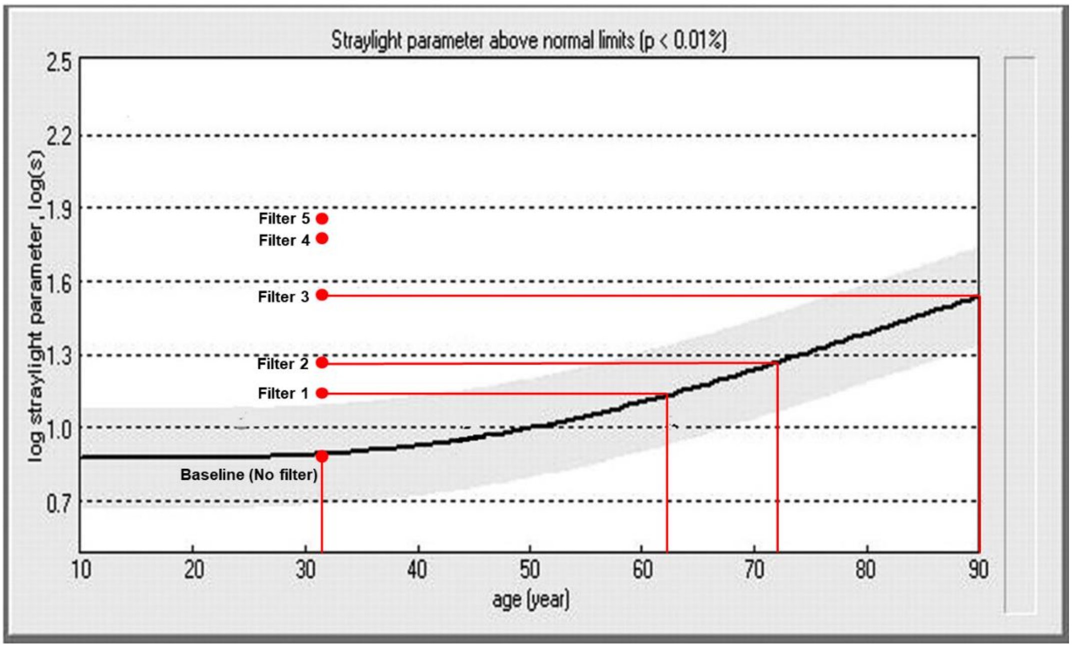


Figure 5.2: Stray light values with increasing age *(adapted from Anderson et al., 2009).* Baseline and subsequent stray light values with Filters 1 to 5 for the 31 year old subject (NS) are also plotted (red circles).

5.2.4 Statistical analysis

The final threshold letter sizes, generated as a percentage with reference to a box size of 512 pixels were converted to a logMAR score for further analysis, where as described in Section 2.1.2; the stroke width for the VO included both the central dark bar and the surrounding white flanks. The GraphPad Prism statistical analysis package (GraphPad Software, Inc., La Jolla, CA) was used to compare VA thresholds using a one-way repeated measures ANOVA, and statistically significant results were investigated using Bonferroni's Multiple Comparison post-hoc analysis for selected pairwise comparisons.

5.3 Results

The average of the three acuity thresholds were plotted against the actual stray light value for each subject for no filter and Filters 1 to 5 (Figure 5.3). Error bars represent the SD of the three threshold measurements for each filter. As expected, and in agreement with previous studies, VA thresholds for the VOs were larger than those attained with conventional letters (Shah et al., 2012, Shah et al., 2011). This was true for all 6 levels of intraocular stray light levels, both in the fovea and the periphery for both subjects. As explained in previous chapters, the visual system must rely on the high SF information only for letter identification with the VOs in which the low SF information is filtered out, resulting in larger acuity thresholds.

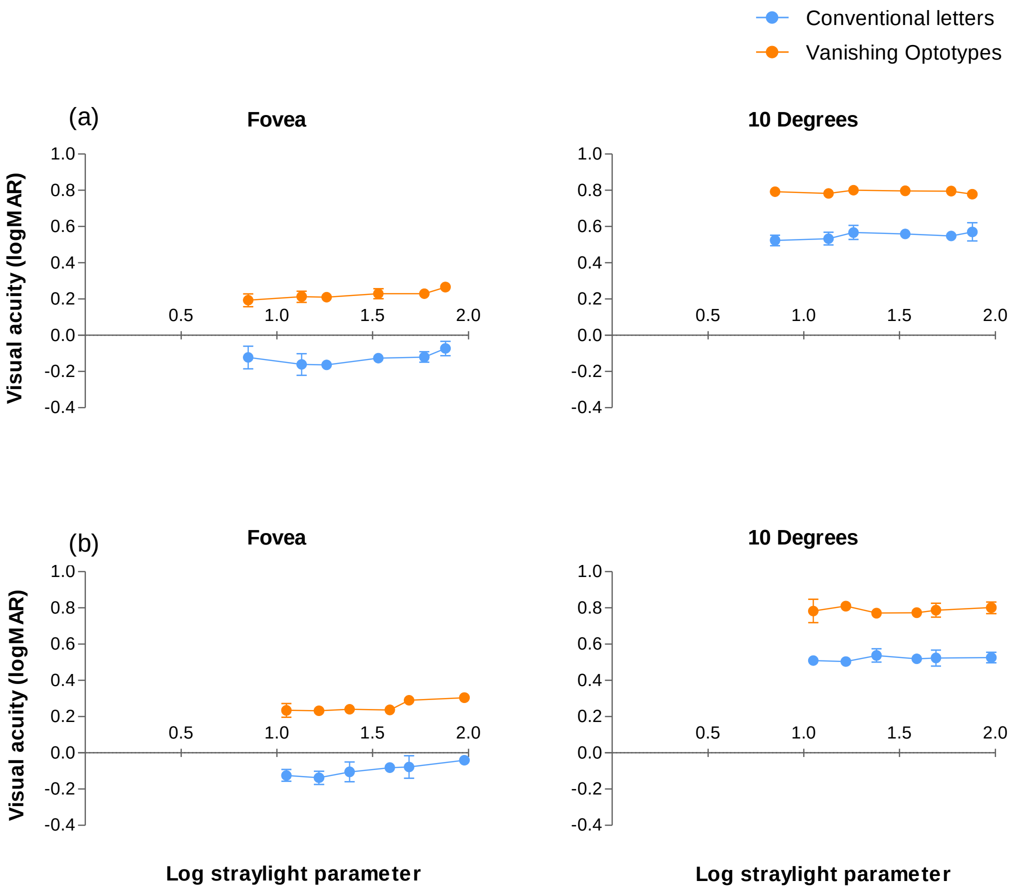


Figure 5.3: Acuity thresholds versus log stray light parameter attained with no filter and Filters 1 to 5. Results are shown for (a) experienced observer NS and (b) naïve subject JM in the fovea and at 10 degrees. Error bars represent the SD of the three threshold measurements for each filter.

In the fovea, it can be seen that acuity thresholds for both subjects were very similar. The range in acuity thresholds for the conventional letters for the six different stray light levels was -0.16 to -0.07 logMAR for subject NS and -0.14 to -0.04 logMAR for subject JM. For the VOs, the range was 0.19 to 0.27 logMAR for subject NS and 0.23 to 0.30 logMAR for subject JM. A one-way repeated measures ANOVA on acuity thresholds yielded significant differences with filters from baseline when measured with the conventional letters $F_{5,1} = 11.62$, $p = 0.0088$ and VOs $F_{5,1} = 10.71$, $p = 0.0105$. However, a post-hoc Bonferroni's Multiple Comparison test indicated that the mean acuity thresholds for the two subjects only became significantly different ($p < 0.05$) from the baseline with the final Filter 5 for conventional letters and with Filters 4 and 5 for the VOs.

Acuity thresholds attained at 10 degrees in the nasal field were again similar for both subjects for the six different stray light levels. The range for the conventional letters was 0.52 to 0.57 logMAR for subject NS and 0.50 to 0.54 logMAR for subject JM. The range in acuity for the VOs was 0.78 to 0.80 logMAR for subject NS and 0.77 to 0.81 logMAR for subject JM. Significant differences in conventional letter thresholds from baseline with different filters was again found with a one-way repeated measures ANOVA, $F_{5,1} = 9.67$, $p = 0.0132$ and investigation with a post-hoc Bonferroni's Multiple Comparison test indicated that the mean acuity thresholds for the two subjects were significantly different ($p <0.05$) from the baseline with Filters 2 and 5. No significant difference from baseline was found for any of the filters with a one-way repeated measures ANOVA, $F_{5,1} = 0.11$, $p = 0.9860$ for the VOs.

5.4 Discussion

In investigating the potential use of VOs in psychophysical tests to detect neural damage specifically, it is important to investigate their performance resistance to optical factors. The previous chapter examined the performance of VOs with optical defocus, finding that recognition acuity thresholds were slightly more robust to the effects of blur compared to conventional letters in the fovea.

Recognition thresholds were also more resistant to optical defocus in the periphery compared to the fovea (Shah et al., 2012a). The rich SF spectra of conventional letters make them particularly vulnerable to the effects of phase reversals caused by optical defocus (Thorn and Schwartz, 1990). Furthermore, whilst recognition thresholds for VO's in the fovea are limited by the low pass filtering effects of the eye's optics, under peripheral viewing conditions, recognition and detection thresholds were found to separate and remained so with even up to +7D of optical defocus. This suggests that, under peripheral viewing conditions, recognition is neural sampling and not optically limited for these targets since the lower density sampling array of the peripheral retina means that higher SFs are neurally unresolvable even if well focused optically (Shah et al., 2012a). The aim of this study was to investigate the effect of increasing wide-angle light scatter on VO thresholds which is particularly relevant in an ageing eye and in those with lenticular opacities. De Wit et al., (2006) have demonstrated the validity of simulating cataract induced stray light by placing a scattering filter in front of the eye.

A mean baseline measure of intraocular stray light of 0.95 log[s] was found for the two subjects, which agrees with other studies with similar aged participants (Michael et al., 2009). Figure 5.3 demonstrates that whilst statistically significant differences in threshold acuity were found with the densest Filter 5 compared to baseline in the fovea for the high contrast conventional letters, this was an actual reduction of only 0.06 logMAR (averaged for both subjects). Similarly, for the VOs, a statistical difference was found with Filters 4 and 5 compared to baseline in the fovea and this was an actual reduction of 0.05 and 0.07 logMAR respectively, i.e. less than one line on a logMAR chart. Furthermore, Figure 5.2 demonstrates the significant density of opacity that Filters 4 and 5 actually represent. In the periphery, thresholds with the VOs were found to be unaffected by even the densest filters and for the conventional letters, whilst differences were found with Filters 2 and 5 from baseline, again this represented a deterioration of only 0.03 logMAR at most. Thus acuity thresholds attained with either the conventional letters or VOs, both in the fovea and the periphery, appear to be robust to the effects of simulated lens opacification inducing wide-angle light scatter.

This study simulated wide-angle scatter only and is not intended to predict the average performance for all cataracts which vary in colour, location, shape as well as density within the lens (Chua et al., 2004). However the results suggest that, for the conditions in this study, only substantial levels of wide angle light scatter corresponding to levels found on average in those over 90 years old or with significant cataract, reduce the contrast and/or retinal illuminance significantly enough to affect conventional letter and VO recognition performance. This is in line with reports which suggest that high contrast conventional black-on-white letter recognition acuity is a poor indicator of the functional impairment which may be caused by light scatter (Amesbury et al., 2009, Van Den Berg et al., 2007, Elliott et al., 1990, Zhu et al., 2016, Elliott et al., 1996).

The initial studies in this thesis have demonstrated that recognition acuity with high-pass letters is likely to be sampling limited outside the fovea and, furthermore, generally robust to the effects of optical defocus (Shah et al., 2012a) and intraocular light scatter. Thus, the role of high-pass letters in a recognition acuity task in separating neural from optical losses of vision appears to be promising. The subsequent chapters will explore how these can be incorporated into a test which has the potential to be used in a clinical application.

6. Design of a novel high-pass letter acuity chart - effect of scoring and termination rules

CHAPTER 6

The work discussed within this chapter has been published in a peer reviewed journal;

Shah N., Dakin S. C., Whitaker H. L. & Anderson R. S. Effect of scoring and termination rules on test-retest variability of a novel high-pass letter acuity chart. Investigative Ophthalmology and Visual Science. 2014;55:1386-92.

6.1 Introduction

VA measurement is recognised as being the most universally employed assessment of visual function in the detection and monitoring of refractive error and/or ocular disease and when evaluating treatment efficacy. Furthermore, it is also often the standard outcome measure reported in numerous studies and clinical trials (Bai et al., 2011, Daien et al., 2012, Goh et al., 2011, Gonnermann et al., 2012, Shona et al., 2011).

The currently considered gold standard tests of VA, logMAR charts and in particular the ETDRS chart, have greatly improved the accuracy of measurements compared to the Snellen chart (Lovie-Kitchin, 1988, Kaiser, 2009, Falkenstein et al., 2008, Lim et al., 2010) for reasons explained earlier in Section 1.2. However, it is important to note that whilst detection of optical inadequacies within the visual system is excellent with the conventional black-on-white letter designs employed in these charts (Thorn and Schwartz, 1990, Nestares et al., 2003, Ravikumar et al., 2010), neural deficiencies may not be identified as effectively (Klein et al., 1995, Wong et al., 2008).

Secondly, owing to the more uniform test design, TRV which limits a test's specificity and sensitivity to genuine changes in clinical status (see Section 1.3.2), are much improved by employing logMAR rather than Snellen charts (Rosser et al., 2001, Laidlaw et al., 2003, Lim et al., 2010). Section 1.3.3 reviewed additional factors beyond the test chart design which can influence TRV, some of which are potentially controllable such as uncorrected refractive error (Rosser et al., 2004, Carkeet et al., 2001) and others not so, such as the presence of ocular disease (Patel et al., 2008). Investigations have also suggested that both VA scores and their reliability can be affected by the number of alternative letter choices available (Shah et al., 2011a, Carkeet, 2001) or the assumptions made by the subject about this (McMonnies, 1999). A modification in scoring methods can improve TRV (Carkeet, 2001, Arditi and Cagenello, 1993) with a single-letter scoring system (see Section 1.3.3), which offers a finer gradation of measurement (Bailey et al., 1991), being discernibly superior to the line-assignment scoring method. Laidlaw et al., (2003) demonstrated an improvement in TRV scores with a line-assignment scoring method from +/- 0.30 to +/- 0.20 logMAR by using the ETDRS rather than Snellen chart and a further improvement in TRV to +/- 0.14 logMAR for the ETDRS chart by employing a single-letter scoring algorithm. Bailey and Lovie (1991) also reported an improvement in logMAR chart TRV from +/-0.20 to +/-0.10 logMAR by using letter rather than line scoring. Inter-examiner variability can also have a detrimental influence on TRV values (Gibson and Sanderson, 1980) and the implementation of forced choice testing procedures using pre-set test termination criteria can be beneficial in attempting to negate this. Carkeet (2001) employed exact calculation and Monte Carlo simulation to investigate the effect of two different sets of termination rules on the mean and SD of logMAR scores attained with different nAFC. The two termination rules used in these investigations were a 'line-by-line' termination rule in which a set number of mistakes on a line are made and a 'letter-by-letter' termination rule in which a set number of mistakes across the whole chart are made. The concluding recommendations from this study were that a termination rule of four mistakes or more on a single line should be implemented for the ETDRS chart using single-letter scoring. Arditi and Cagenello (1993) suggest an absolute upper limit of reliability of +/-0.10 logMAR for the ETDRS chart under carefully controlled optimal testing conditions in practiced observers with good VAs. Under routine clinical testing conditions, values would be expected to be larger than this. Indeed, TRV values of up to 2 logMAR lines have been reported (Hazel and Elliott, 2002, Vanden Bosch and Wall, 1997, Elliott and Sheridan, 1988, Lovie-Kitchin, 1988, Reeves et al., 1991, Brown and Lovie-Kitchin, 1993, Bailey et al., 1991, Rosser et al., 2003a, Manny et al., 2003).

Significant differences were revealed in the individual recognition thresholds for each letter of the alphabet in conventional letter design in Chapter 4. The legibility range reduced from 0.40 logMAR when considering the whole alphabet to 0.22 logMAR for the Sloan letter set only (Shah et al., 2011a). In particular, the circular letters (C G O Q) were found to behave as a separate subset displaying the highest recognition thresholds (Shah et al., 2011a). Whilst the Sloan letters are arranged in combinations on the ETDRS chart such that each line is of similar difficulty (Ferris et al., 1982), it is the inter-letter legibility differences within a line which are thought to contribute significantly to acuity measurement variability (McMonnies and Ho, 2000). Thus, by selecting an alternative set of 10 conventional design letters, with more similar recognition thresholds and eliminating the circular letters, it may be feasible to further reduce TRV. Furthermore, the investigations in Chapter 3 demonstrate better measurement repeatability and VA thresholds which are less affected by the nAFC and the actual letter choices involved (Shah et al., 2011a) by removing the low SF content where conventional design letters particularly differ (Anderson and Thibos, 1999a, Anderson and Thibos, 1999b, Gervais et al., 1984).

Thus the aim of this study was to investigate the effects that different scoring methods and test termination criteria have on VA thresholds and TRV in normal subjects with a range of acuities owing to uncorrected refractive error, with:

a) Charts employing a new alternative 10 letter alphabet set with more similar recognition thresholds compared to the Sloan letter set.
b) New VA charts incorporating high-pass filtered letters compared to conventional letter designs.

6.2 Methods

6.2.1 Visual acuity charts

A total of eight different letter acuity charts were graphically designed and digitally produced by a reprographics professional at Ulster University, Coleraine, N. Ireland. The first two charts were constructed with the conventional 5 x 5 matrix, black-on-white, letter design using the 10 Sloan letters (C D H K N O R S V Z). The letter combinations on these charts, Conventional Sloan letter set 1 and 2 (C S1 and C S2) were identical to ETDRS charts 1 and 2 respectively. The third and fourth charts were constructed using the same Sloan letter combinations in high-pass rather than conventional letter design with a grey background of the same mean luminance as the letters and these were named the Moorfields Acuity Chart Sloan letter set 1 and 2 (MAC S1 and MAC S2). The luminance of the white components of each chart was 114 cd/m^2 and that of the grey background of the high-pass charts was 50.5 cd/m^2. The letter size ranged from 1.20 to -0.20 logMAR 'equivalent' (based on stroke width where stroke width for the VO letters consisted of the dark middle bar and its two white flanks) from a 4 m test distance. Considering that larger VA thresholds had been demonstrated with VOs compared to conventional letters (Shah et al., 2011a, Shah et al., 2012a) in the previous investigations of Chapters 3 to 5, an extended range was used rather than the range typically found on the ETDRS charts (1.0 to -0.30 logMAR) in order to avoid the potential issue of comparing results between charts used at different test distances (Dong et al., 2002). The new chart constructions otherwise replicated that of the conventional ETDRS chart layout, with letters spaced a letter width apart on each line and lines spaced a letter height apart from the line below. Each line was composed of 5 letters.

The remaining four charts used a different 10-letter alphabet (B E H K N P R S X Z). The letters incorporated in this chart were selected based on the findings in Chapter 4 (Shah et al., 2012a) as follows:

a) The circular letters (C G O Q) were first excluded since their higher recognition thresholds relative to the other letters suggested that they seemed to behave as a separate subset of their own.
b) Those letters which displayed the greatest differences in their recognition thresholds in conventional letter format (A C G J L O Q U) were also eliminated.
c) Letters which were deemed either too simple (I J L U V Y) or too complex (M W) in their spatial form were removed.
d) Since an ensuing aim of this work was to design a VA chart better capable of detecting factors affecting the neural structure of the retina, letters which demonstrated the least sampling limited behaviour with the smallest differences in their detection and recognition thresholds under peripheral viewing conditions in high-pass format (A D J L T U V) were also removed. This left the following, B E F H K N P R S X Z from which the letter 'F' was also eliminated to leave ten final letters.

These ten letters (B E H K N P R S X Z) were incorporated in conventional letter format in two different randomised orders, Conventional New letter set 1 and 2

(C N1 and C N2) and in the same order in high-pass design for the Moorfields Acuity Chart New letter set 1 and 2 (MAC N1 and MAC N2). Figure 6.1 displays the designs of the four Conventional and four Moorfields Acuity Charts used in this study.

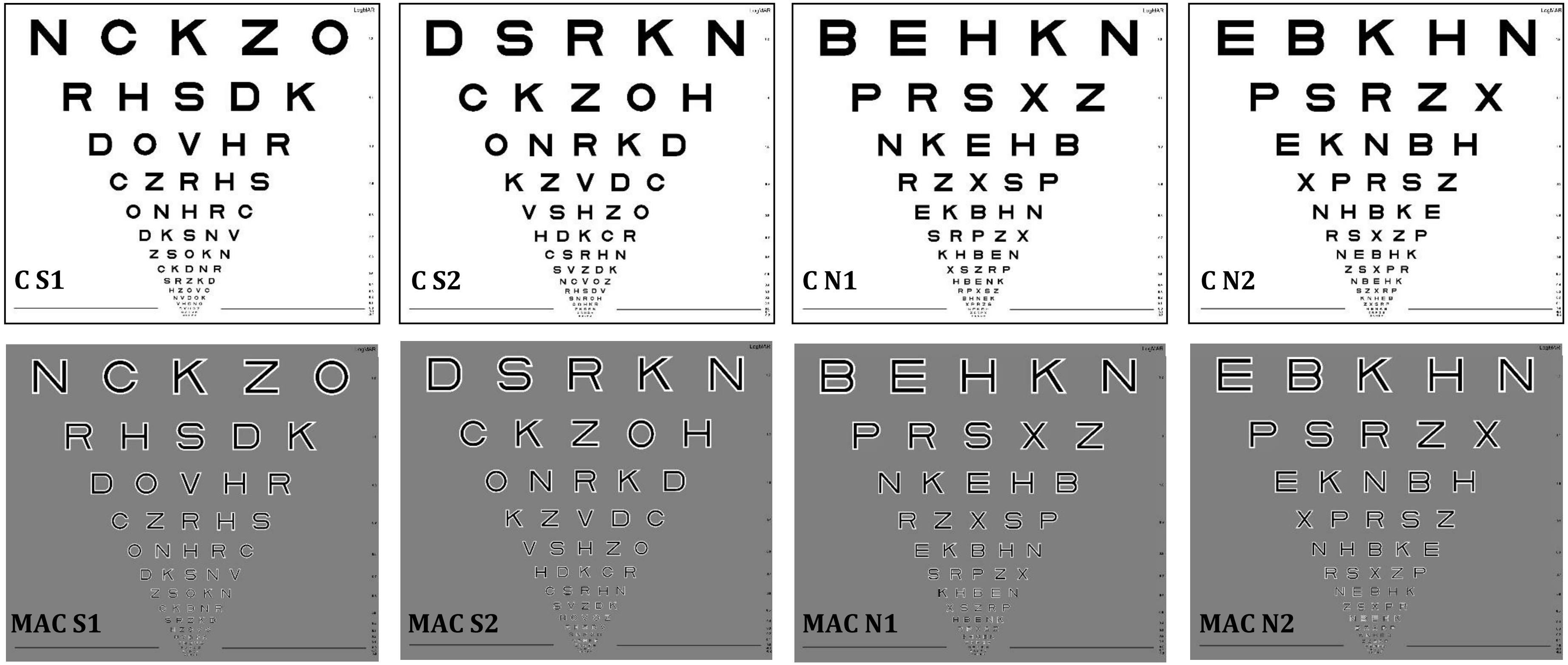


Figure 6.1: The appearance of charts C S1, C S2, MAC S1 and MAC S2 (which use the Sloan alphabet set) and C N1, C N2, MAC N1 and MAC N2 (which use the New alphabet set). The image luminance may not be balanced due to reproduction limitations.

6.2.2 Subjects

Fifty normal observers (17 male participants) ranging in age from 20 to 76 years (mean age 42.8 years) were recruited from a primary care optometric practice (Hynes Optometrists, London, UK) and from the staff of Moorfields Eye Hospital, London, UK. Preliminary tests were conducted in order to determine each participant's eligibility to enrol in the study. The first of these was baseline refraction (retinoscopy and subjective) at 4 m using ETDRS chart R as described in Section 2.2.4 and the mean spherical refractive error for this group was found to be -0.93D (range -5.38 to +3.00D). A recruit was excluded from the study if uncorrected vision fell outside the measurement range of the charts at 4 m. Thus mean unaided visions using chart C S1 for the 50 eligible subjects was 0.44 logMAR (range -0.04 to 1.16 logMAR). In order to exclude any ocular pathology, ocular examination was conducted using slit lamp biomicroscopy and slit lamp binocular indirect ophthalmoscopy. The test eye was randomly assigned for each subject using an online random assignment software (described in Section 2.4) with the right eye tested in 25 subjects and the left eye in the other half. Ethical approval was obtained for this study from the West London Research Ethics Committee and all procedures adhered to the tenets of the Declaration of Helsinki.

6.2.3 Procedure

Unaided vision was measured in the test eye of each subject using the eight different acuity charts in a randomised order in a single visit, determined again using an online randomiser (described in Section 2.4) in order to control for fatigue and learning effects. Subjects were asked to identify each letter starting from the top left of the chart and were encouraged to guess when unsure since this has been shown to improve VA scores (Smith, 2005). Viewing time was kept unrestricted since the investigations of Heinrich et al., (2010) demonstrated better acuity thresholds and TRV with increasing exposure duration up to their maximum tested 10 s. The response for each letter was recorded by the examiner on a pro forma data sheet with a circle drawn around each correctly identified letter and a cross through each incorrect letter. The test was terminated when a whole line was incorrectly identified so that VA scores could be determined retrospectively in three different ways using various scoring methods and termination criteria (Carkeet, 2001, Vanden Bosch and Wall, 1997):

a) *Letter-by-letter scoring with line-based termination* whereby each letter was assigned a value of 0.02 logMAR and the test terminated with a specific number of errors per line i.e. one-or-more, two-or-more, three-or-more, four-or-more or five (whole line) wrong per line. The subject was allowed to complete the line being attempted even if the termination criteria had already been met as per conventional practice and the acuity was calculated in logMAR terms using Equation 2.1:

   VA_{4m} (logMAR) = (1.2 + 0.1) – C_N*0.02.

b) *Letter-by-letter scoring with chart-based termination* whereby each letter was assigned a score of 0.02 logMAR and the test terminated with a specific number of errors across the whole chart, i.e. one-or-more, two-or-more, three-or-more, four-or-more or five-or-more letters wrong across the chart. Simulations by Carkeet (2001) revealed similar results with the subject completing the line after a criterion number of mistakes compared to when the line was not completed for letter-by-letter scoring with whole-chart termination, although the latter gave slightly higher acuity scores and slope-corrected SDs. Thus, for this study, the subject was again allowed to complete the line being attempted even if the termination criteria had already been met and the acuity was calculated in logMAR using Equation 2.1: VA_{4m} (logMAR) = (1.2 + 0.1) – C_N*0.02.

c) *Line-by-line scoring* where the VA score was taken as the last line in which at least 3 out of 5 letters were read correctly.

The scoring method employed in the first two techniques is the same but they differ in their termination criteria. The third technique uses a different scoring method.

6.2.4 Statistical analysis

The difference in the two acuity measurements with each chart type (C S1 and C S2, MAC S1 and MAC S2, C N1 and C N2 and lastly MAC N1 and MAC N2) were calculated for each individual for each of the different scoring and termination rules. The Shapiro-Wilk W-test (and frequency distribution plots, where appropriate) were used to confirm that these differences were normally distributed and the methods of Bland and Altman (Bland and Altman, 1986) and ordinary least squares regression analysis were employed. The GraphPad Prism statistical analysis package (GraphPad Software, Inc., La Jolla, CA) was used for these purposes.

6.3 Results

The mean unaided vision results for C S1 with letter-by-letter scoring for line- and chart-termination of 5 letters wrong and line-by-line scoring were respectively found to be 0.44 logMAR (range -0.04 to 1.16 logMAR), 0.46 logMAR (range -0.04 to 1.20 logMAR) and 0.44 logMAR(range -0.10 to 1.20 logMAR). The corresponding values for MAC S1 were 0.57 logMAR (range 0.08 to 1.14 logMAR), 0.58 logMAR (range 0.08 to 1.14 logMAR) and 0.56 logMAR (range 0.00 to 1.2 logMAR).

The Shapiro-Wilk W test was used to test the null hypothesis that the differences in acuity found between charts 1 and 2 of each type were sampled from a normal distribution. This was not rejected for letter-by-letter scoring line-based termination of five letters with $p = 0.251$ for C S1 and C S2, $p = 0.309$ for MAC S1 and MAC S2, $p = 0.412$ for C N1 and C N2 and $p = 0.286$ for MAC N1 and MAC N2. This was also not rejected for letter-by-letter scoring chart-based termination of five letters with $p = 0.070$ for C S1 and C S2, $p = 0.412$ for MAC S1 and MAC S2, $p = 0.149$ for C N1 and C N2 and $p = 0.286$ for MAC N1 and MAC N2. The null hypothesis was rejected for line-by-line scoring with $p = 0.001$ for C S1 and C S2, $p = <0.0001$ for MAC S1 and MAC S2, $p = 0.004$ for C N1 and C N2 and $p = <0.0001$ for MAC N1 and MAC N2. However, frequency distribution plots to further examine this distribution (Figure 6.2) demonstrated no gross deviation from normal, and on this basis the use of Bland and Altman summary statistics in terms of mean bias and TRV expressed as the 95% LOA were considered justified.

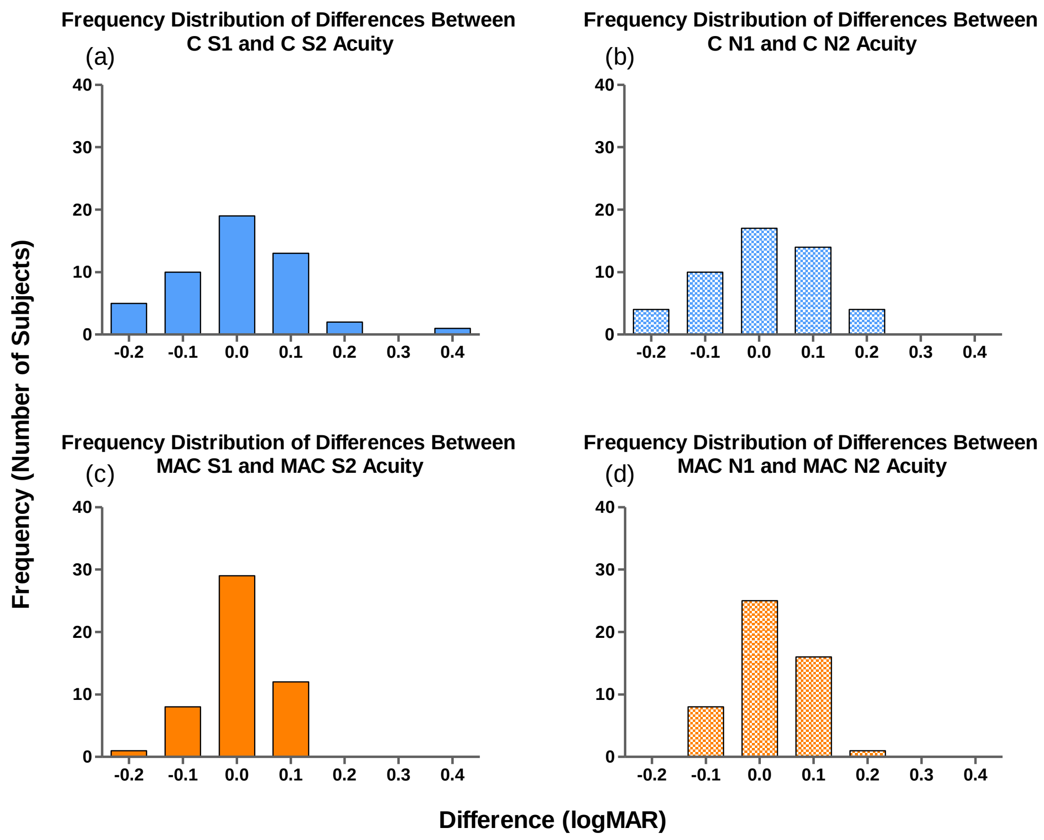


Figure 6.2: Frequency distribution plots of the difference in acuity attained between charts with line-by-line scoring. Results are presented for (a) C S1 and C S2, (b) C N1 and C N2, (c) MAC S1 and MAC S2 and (d) MAC N1 and MAC N2 charts and show no gross deviation from a normal distribution.

Bland and Altman scatter plots, graphically presenting the spread of results were constructed for each of the chart combinations for letter-by-letter scoring with line-based (Figure 6.3) and chart-based (Figure 6.4) termination of five letters wrong and line-by-line scoring (Figure 6.5). In each figure, (a-d) displays the TRV of each letter set and chart type together with the mean bias in scores. The results indicate no systematic association between the underlying acuity and level of agreement between charts 1 and 2 or TRV each time using either scoring technique or termination criteria.

Table 6.1 summarises the Bland-Altman statistics for (a) letter-by-letter scoring with line-based and (b) chart-based termination of five letters wrong, and (c) line-by-line scoring. The null hypothesis that the SD for C S1-C S2 is not significantly different to that of MAC S1-MAC S2 was examined using the F-test and not accepted for any of the scoring or termination methods (two-tailed $p < 0.05$, $F_{49,49} = 1.61$). Indeed, for line-by-line scoring, TRV scores improved by almost a line with the high-pass letter charts compared to the conventional letter charts. No significant difference was found in SDs between the Sloan letter set and the New alphabet set in either conventional or high-pass letter design for any of the scoring or termination methods (two-tailed $p > 0.05$, $F_{49,49} = 1.61$).

In each of Figures 6.3 to 6.5, (e) represents the results of a method comparison study giving information on the agreement between the two different chart types (C S1 and MAC S1). For all scoring and termination methods, not only a systematic, but also proportional bias can be deduced such that a greater level of disagreement between the two chart types is apparent at the better acuity end. Ordinary least squares regression analysis was used to investigate and confirm this potential proportional bias for all scoring and termination methods ($r^2 = 0.217$, $p = 0.001$, $r^2 = 0.315$, $p = <0.0001$, $r^2 = 0.204$, $p = 0.001$ for letter-by-letter scoring with line-based and chart-based termination of five letters wrong, and line-by-line scoring respectively). This means that the difference in acuity between C S1 and MAC S1 was -0.21 to -0.22 logMAR (approximately 2 logMAR lines) at the 0.00 logMAR VA level, compared to a difference of only -0.02 to -0.05

logMAR (approximately half a logMAR line) at the 1.00 logMAR VA level depending on the scoring and termination system used.

(a)

Test	Mean Difference (SE)	95% CI Mean Difference	Range of Observed Differences	TRV, 95% Limits of Agreement
C S1 – C S2	-0.001 (0.010)	-0.020, 0.019	-0.18, 0.12	±0.135
MAC S1 – MAC S2	-0.004 (0.007)	-0.017, 0.010	-0.10, 0.12	±0.095
C N1 – C N2	0.008 (0.010)	-0.012, 0.029	-0.18, 0.16	±0.142
MAC N1 – MAC N2	0.022 (0.008)	0.007, 0.037	-0.08, 0.14	±0.105

(b)

Test	Mean Difference (SE)	95% CI Mean Difference	Range of Observed Differences	TRV, 95% Limits of Agreement
C S1 – C S2	0.005 (0.012)	-0.019, 0.029	-0.26, 0.24	±0.166
MAC S1 – MAC S2	-0.001 (0.007)	-0.015, 0.013	-0.10, 0.12	±0.098
C N1 – C N2	0.008 (0.011)	-0.015, 0.030	-0.18, 0.18	±0.154
MAC N1 – MAC N2	0.021 (0.008)	0.006, 0.036	-0.08, 0.14	±0.104

(c)

Test	Mean Difference (SE)	95% CI Mean Difference	Range of Observed Differences	TRV, 95% Limits of Agreement
C S1 – C S2	0.002 (0.017)	-0.031, 0.035	-0.20, 0.40	±0.229
MAC S1 – MAC S2	0.004 (0.010)	-0.016, 0.024	-0.20, 0.10	±0.137
C N1 – C N2	0.002 (0.016)	-0.031, 0.035	-0.30, 0.20	±0.226
MAC N1 – MAC N2	0.020 (0.010)	-0.001, 0.041	-0.10, 0.20	±0.143

Table 6.1: Bland-Altman summary statistics for different scoring and termination criteria. Results are shown for (a) letter-by-letter scoring with line-based termination of five letters, (b) letter-by-letter scoring with chart-based termination of five letters and (c) line-by-line scoring. These data are presented graphically in Figures 6.3 to 6.5.

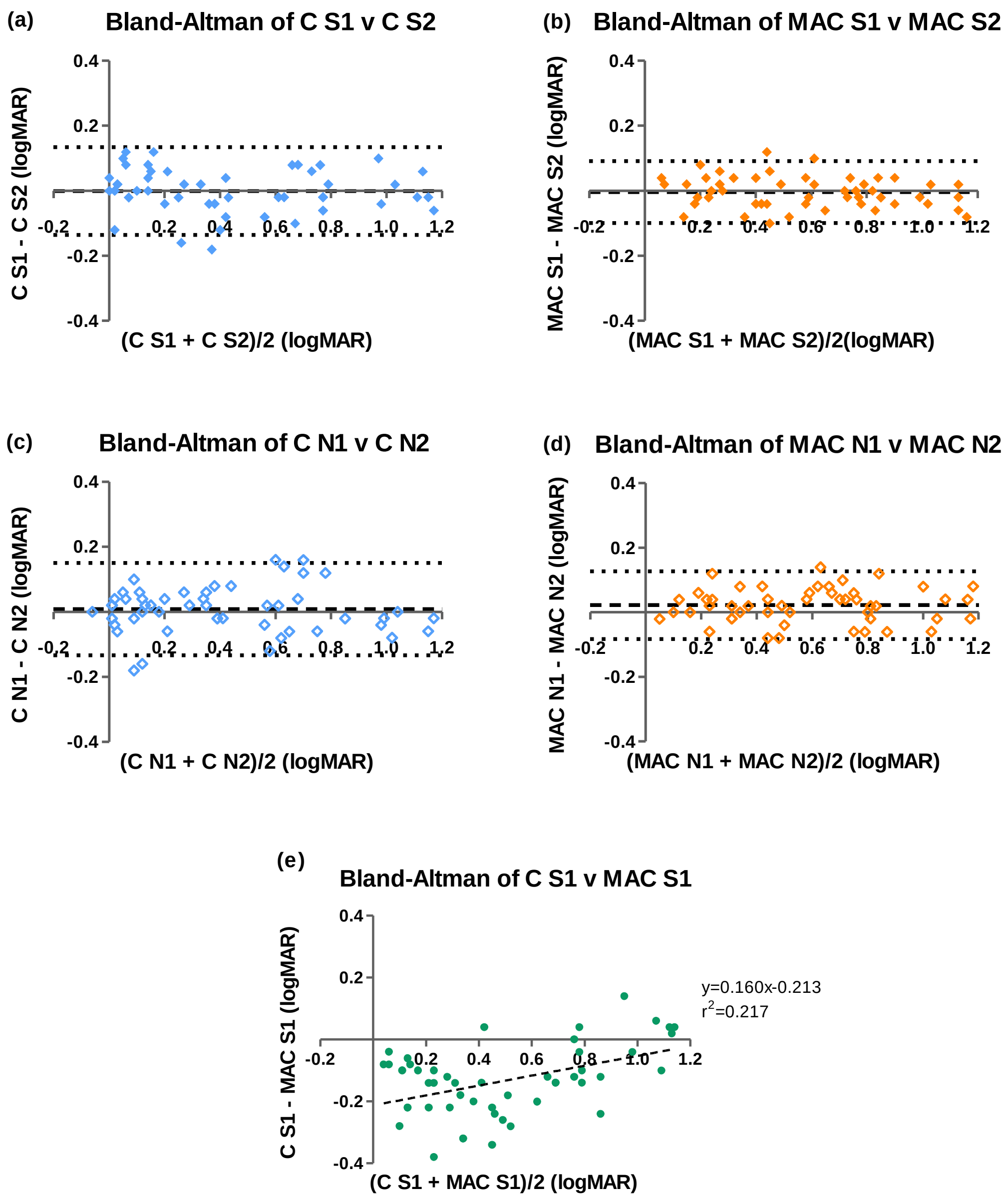


Figure 6.3: Bland-Altman plots for test and retest VA measurements using letter-by-letter scoring with line-based termination of five letters wrong *(adapted from Shah et al., 2014).* The mean difference and upper and lower 95% LOA are plotted for (a) C S1 versus C S2, (b) MAC S1 versus MAC S2, (c) C N1 versus C N2 and (d) MAC N1 versus MAC N2. The Bland-Altman plot for the method comparison study between C S1 and MAC S1 is shown in (e) with the slope fitted using ordinary least squares regression analysis p = 0.001.

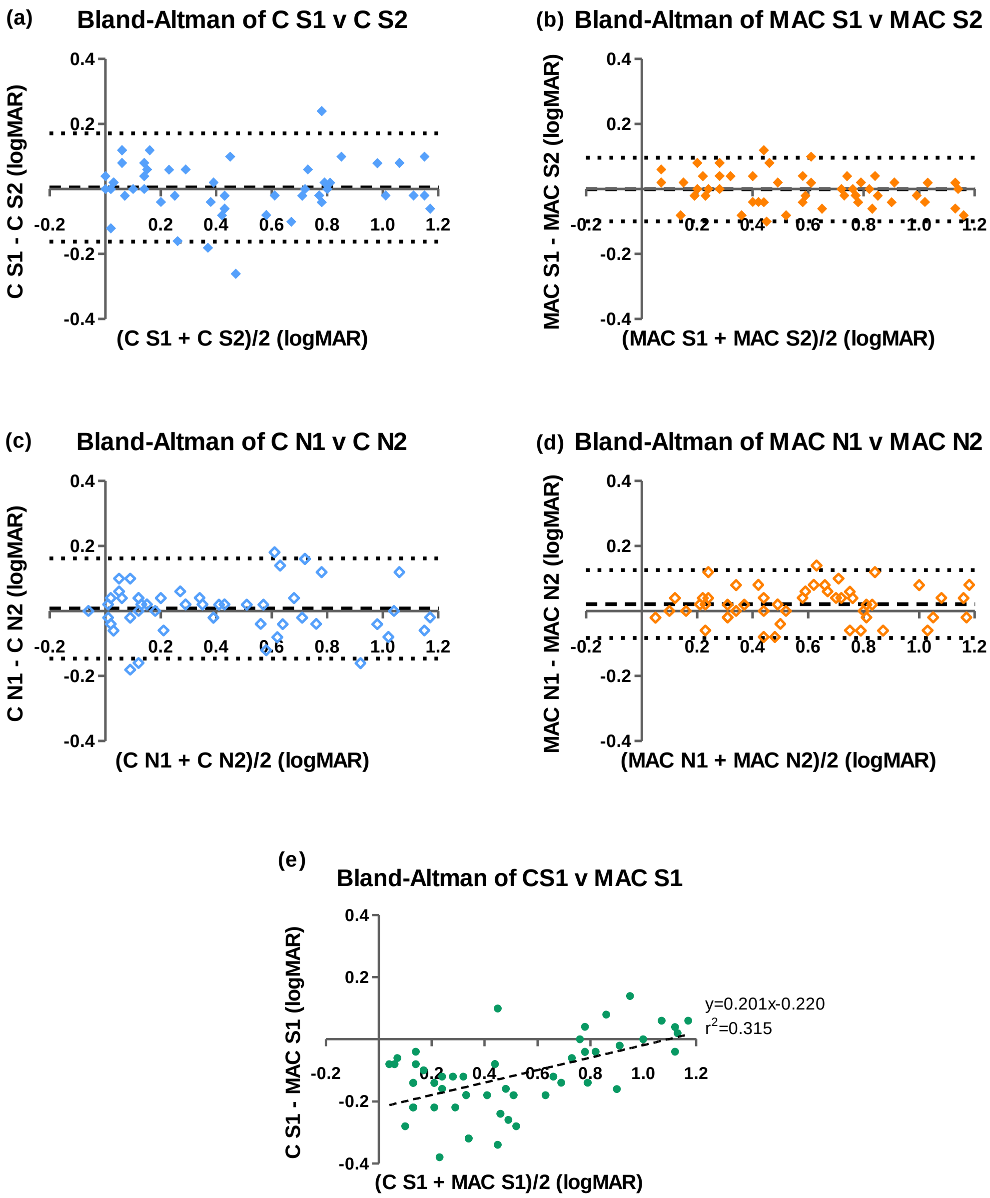


Figure 6.4: Bland-Altman plots for test and retest VA measurements using letter-by-letter scoring with chart-based termination of five letters wrong. The mean difference and upper and lower 95% LOA are plotted for (a) C S1 versus C S2, (b) MAC S1 versus MAC S2, (c) C N1 versus C N2 and (d) MAC N1 versus MAC N2. The Bland-Altman plot for the method comparison study between C S1 and MAC S1 is shown in (e) with the slope fitted using ordinary least squares regression analysis p = <0.0001.

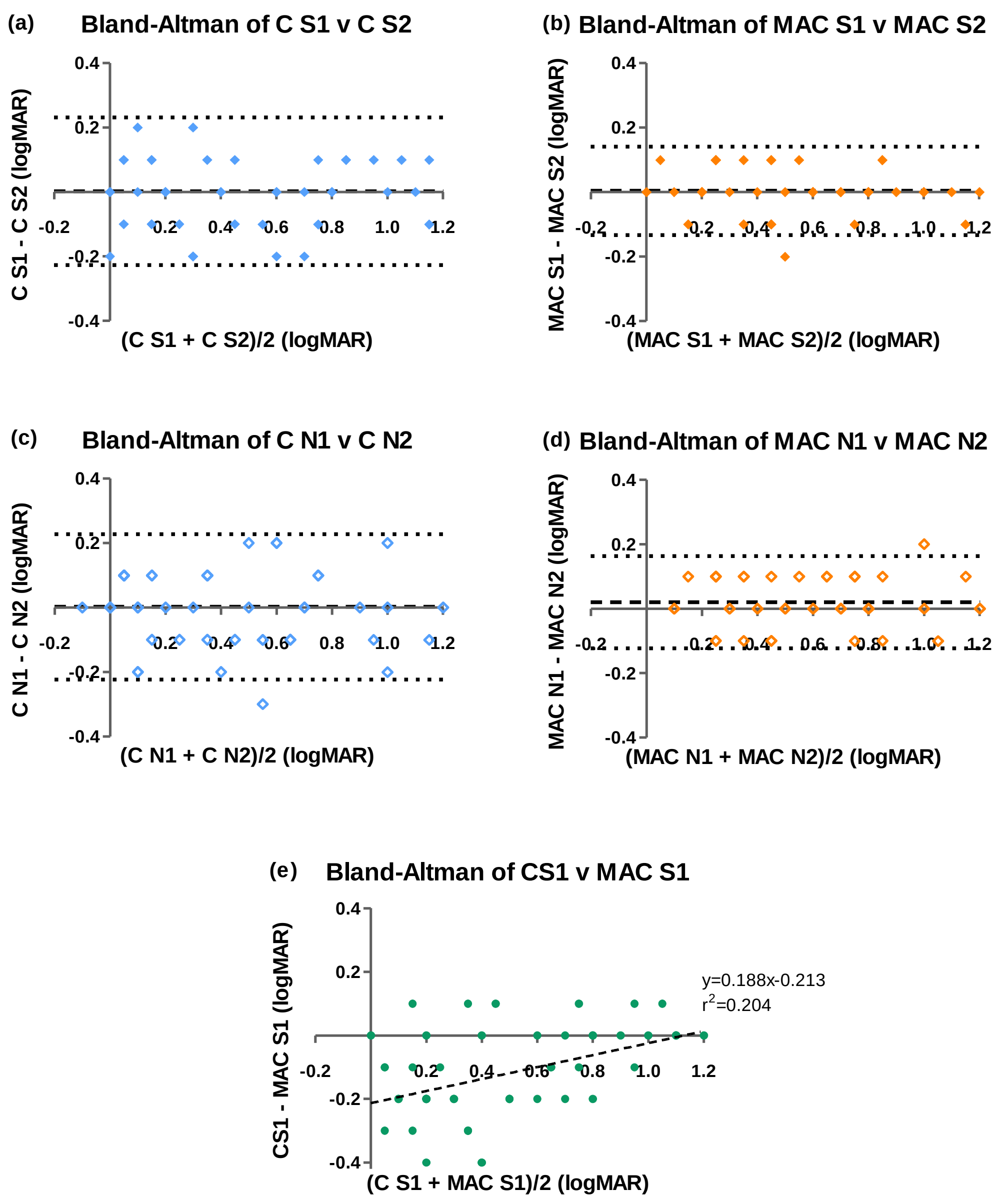


Figure 6.5: Bland-Altman plots for test and retest VA measurements using line-by-line scoring. The mean difference and upper and lower 95% LOA are plotted for (a) C S1 versus C S2, (b) MAC S1 versus MAC S2, (c) C N1 versus C N2 and (d) MAC N1 versus MAC N2. The Bland-Altman plot for the method comparison study between C S1 and MAC S1 is shown in (e) with the slope fitted using ordinary least squares regression analysis p = 0.001.

Figure 6.6 (a and b) illustrates how TRV altered with different line- and chart-based termination rules with letter-by-letter scoring. TRV for C S1-C S2 (Figure 6.6a) was +/-0.14 logMAR for termination of five and four-or-more letters wrong per line, but with a termination criterion of three-or-more letters wrong per line this increased to +/-0.17 logMAR. These TRV values are comparatively higher than for the MACs. TRV for MAC S1-MAC S2 remained at +/-0.10 logMAR for five, four- and three-or-more letters wrong per line and increased to only +/-0.13 logMAR with a termination criteria of two-or-more letters wrong per line. No benefit to TRV was found using the New alphabet set in either conventional or high-pass letter design with line-based termination criteria.

TRV for C S1-C S2 was higher for chart-based termination than with line-based termination rules (Figure 6.6b). TRV was +/-0.17 logMAR with five-or-more and four-or-more errors across the chart and increased slowly thereafter. TRV with chart-based termination with five-or-more letters wrong across the whole chart for MAC S1-MAC S2 (+/-0.10 logMAR) was similar to that for line-based termination. This increased by half a letter for four-or-more errors and again by half a letter for three-or-more errors across the chart. Whilst no benefit was found to TRV of using the New alphabet set in high-pass letter format with chart-based termination rules, a small improvement of one letter was found with the New alphabet set compared to the Sloan letter set in conventional letter format for five-, four- and three-or-more errors across the chart.

As mentioned earlier, TRV for line-by-line scoring is summarised in Table 6.1c. TRV was found to increase more for the conventional letter charts than the MACs with line-by-line scoring compared to letter-by-letter scoring with values of +/-0.23 logMAR for the conventional charts and +/-0.14 for the MACs.

Figure 6.6 (c and d) present the mean logMAR VA scores and how these changed with different line- and chart-based termination rules with letter-by-letter scoring. With line-based termination (Figure 6.6c), the mean VA score for C S1 was 0.44 logMAR for five letters wrong per line. This increased by half a letter for a termination rule of four-or-more letters wrong and by half a letter again for three-or-more letters wrong before jumping to a mean VA threshold of 0.49 logMAR with a termination rule of two-or-more letters wrong per line. Whilst the mean VA logMAR score with MAC S1 was larger for all termination rules than C S1, mean VA was 0.57 logMAR for five letters wrong per line and changed little until a termination criterion of one-or-more letters wrong per line when the mean threshold increased to 0.63 logMAR.

A similar pattern was observed for chart-based termination rules (Figure 6.6d) with larger mean threshold scores for MAC S1 compared to C S1 for all termination criteria. With chart-based termination, the mean VA score for C S1 was 0.46 logMAR for five-or-more letters wrong across the whole chart and this increased by half a letter again for a termination rule of four-or-more letters wrong and by half a letter again for three-or-more letters wrong before jumping to a mean VA threshold of 0.51 logMAR for two-or-more letters wrong and 0.55 for one-or-more letters wrong. Once again, MAC S1 acuity changed little from 0.58 logMAR with five-or-more letters wrong across the chart until a termination criterion of one-or-more letters wrong across the chart when the threshold was 0.63 logMAR.

Mean VA for line-by-line scoring (not plotted) was 0.44 logMAR for C S1, 0.41 logMAR for C N1, 0.56 logMAR for MAC S1 and 0.59 logMAR for MAC N1.

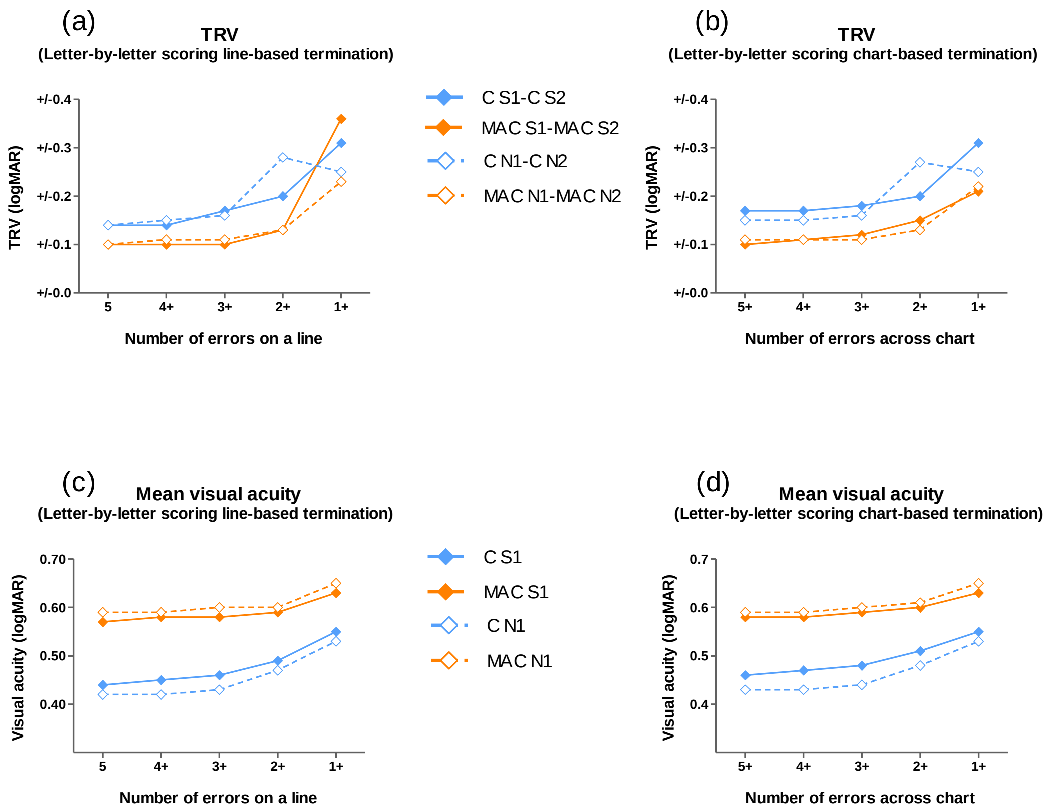


Figure 6.6: TRV (top) and mean VA (bottom), for letter-by-letter scoring with (a and c) line- and (b and d) chart-based termination, with different numbers of letters wrong *(adapted from Shah et al., 2014)***.**

6.4 Discussion

In order to effectively detect small changes in clinical status, it is important that VA tests are accurate and reliable. LogMAR charts have demonstrated their superiority over Snellen charts in this respect, attributable to better standardised design features (Bailey and Lovie, 1976, Ferris et al., 1982). Yet thresholds obtained can still be very dependent on the way in which the test is administered and Carkeet (2001) used simulations to demonstrate the effect of termination rules and nAFC on the final mean and SD of logMAR VA scores. The work of the previous chapters in this thesis (Shah et al., 2011a, Shah et al., 2012a) suggested that it may be appropriate to create a chart using high-pass filtered letters (the Moorfields Acuity Chart) to obtain measurement scores which are less affected by these factors. The aim of this study was to measure and compare actual human performance with conventional letter charts and charts using high-pass filtered letters to explore the effect of scoring and termination rules on acuity.

Two different scoring methods (letter-by-letter and line-by-line) and two different test termination criteria (line-based and chart-based) were assessed and regardless of which was used, the MACs incorporating high-pass letters, demonstrated lower TRV values than the conventional letter logMAR charts in uncorrected normal subjects. In addition, TRV values for the conventional letter charts were also found to be dependent on the specific termination and scoring method used (+/-0.14 logMAR at best for line-based termination and +/-0.17 logMAR at best for chart-based termination with letter-by-letter scoring and +/-0.23 logMAR for line-by-line scoring). This is in agreement with the findings of Raasch et al., (1998) that the standard error of VA measurements yield values which are approximately 1.6 to 1.7 times larger using whole line scoring methods than individual letter scoring.

Figure 6.3 demonstrates how the lowest TRV for the conventional letter charts were found with line-by-line termination criteria increasing from +/-0.14 logMAR for both four-or-more and five letters wrong per line to +/-0.17 logMAR with a termination of three-or-more letters wrong. The outcomes of this current study therefore also support the recommendations of Carkeet (2001) of using a line-by-line termination rule of four or five letters wrong per line for the ETDRS chart with single letter scoring. The TRV values calculated from the data in this study for the conventional letter design charts are typical of those reported in other studies for logMAR acuity measurements with TRV values of up to 0.2 logMAR reported in subjects with unchanged acuity (Hazel and Elliott, 2002, Arditi and Cagenello, 1993, Vanden Bosch and Wall, 1997, Elliott and Sheridan, 1988, Lovie-Kitchin, 1988, Reeves et al., 1991, Brown and Lovie-Kitchin, 1993, Bailey et al., 1991, Rosser et al., 2003a, Manny et al., 2003).

TRV for the high-pass charts however was more similar across all termination and scoring methods (+/-0.10 logMAR at best for both line- and chart-based termination with letter-by-letter scoring and +/-0.14 logMAR for line-by-line scoring) and was also less affected by the choice in number of errors termination criteria. Whilst these differences in TRV between the MACs and conventional letter charts may seem modest (for example, +/-0.10 vs. +/-0.14 logMAR for letter-by-letter scoring with termination of five letters wrong on a line), they represent almost a 29% reduction in the confidence limits for significant change as described by Cousens et al., (2004). Furthermore, these are comparable to the improvements in TRV reported by studies comparing logMAR and Snellen charts (Lim et al., 2010) or different termination criteria (Carkeet, 2001, Arditi and Cagenello, 1993, Laidlaw et al., 2003).

The introduction to this thesis (Section 1.4) explored in detail the selection of different letter choices on VA charts. Whilst the ten Sloan letters are arranged in combinations on the ETDRS chart such that the line average difficulty scores are similar (Ferris et al., 1982), Arditi and Cagenello (1993) calculated the effect that random combinations of five out of the ten Sloan letters on each line would have on variability and found this to be negligible. The significant difference in letter legibility within a line (emphasised in Figure 1.5) is thought to contribute to TRV (Alexander et al., 1997) and Raasch et al., (1998) demonstrated that score variance can mostly be attributed to letters near threshold (demonstrated in Figure 1.8). Whilst a rational potential solution to improve TRV would therefore be to select letters of more similar discriminability (Bailey and Lovie, 1976, McMonnies, 2003) the choice of these can be challenging since legibility can vary with individual observer bias (Bennett, 1965) determined, for example, by previous experiences and cultural backgrounds. Raasch et al., (1998) found minimal difference between variability scores attained with the Sloan letter set compared to a new letter set in which the probability of identification curves was supposedly identical for all the letters and concluded that the inequality of identifiability of the Sloan letters is not in fact a significant factor in TRV. The results of the current study concur with this such that minimal differences in TRV were found between the Sloan letter set and the New alternative alphabet set in either conventional or high-pass filtered letter design. The improvement in TRV observed thus appear to arise mostly from using a high-pass letter design rather than a different alternative ten letter alphabet set. Previous investigation (Chapter 4) has demonstrated a greater resistance in the performance of high-pass letters to optical defocus (Shah et al., 2012a) which is known to increase TRV values for conventional letter charts (Rosser et al., 2004, Carkeet et al., 2001). This may account for the lower values observed for the MAC charts in this study.

A systematic bias was observed between C S1 and MAC S1 measurements (Figure 6.3e, Figure 6.4e and Figure 6.5e) in line with previous findings (Chapters 3 to 5) that acuity thresholds with high-pass filtered letters are larger than those with conventional letters (Shah et al., 2011, Shah et al., 2012) since the visual system must rely on finer features in the spatial domain or higher letter frequencies in the frequency domain (compared to the same logMAR value conventional letter) to be correctly resolved. Acuity values are known to vary between charts and this can depend on the targets employed such as gratings (Anderson and Thibos, 1999a), Landolt rings (Plainis et al., 2013, Becker et al., 2011) and pictures (Shah et al., 2012b, Geer and Westall, 1996), the design and configuration of the chart (Kaiser, 2009, McMonnies, 1999), the number of alternatives (Shah et al., 2011a, Carkeet, 2001) and the choice of letter subset (Shah et al., 2011a, Anderson and Thibos, 2004) depending on which features or SFs are made available to the visual system to resolve. However, ordinary least squares regression analysis of the Bland-Altman plots in this current study confirmed a proportional bias also, such that the difference in VA between C S1 and MAC S1 increased the better the acuity. In acuity chart comparisons, finding a proportional bias is not uncommon. Indeed both Kaiser (2009) and Falkenstein et al., (2008) demonstrated an increasing difference between the ETDRS and Snellen charts with worsening acuity. Since the letter design in both of these charts is the same, the reason for the bias must be attributable to the chart design and layout. In this current study, the bias must be attributable to the letter designs since the chart configurations are identical. Acuity with the high-pass letters is less affected by uncorrected refractive blur as seen by the smaller range in acuities attained with the MAC charts compared to the conventional charts and accounts for the proportional bias observed between the charts. Thus, those with poor acuity attributable to significant uncorrected refractive error can be expected to attain a score half a logMAR line better with the conventional charts compared to the MACs whilst those with better acuities due to only small uncorrected refractive errors will achieve a score approximately two logMAR lines better.

Thus, in conclusion, VA measurements attained with the new MAC employing high-pass letters seem to be less affected by scoring methods and termination rules than an ETDRS style chart employing conventional letters. This is important in routine clinical practice where strict testing procedures may not be rigidly adhered to as perhaps they would for a clinical trial and would also be beneficial in comparing acuity scores attained in different clinical locations or by different clinicians within the same location if identical testing protocols are not applied. Importantly also, the MACs display lower TRV with uncorrected refractive error than the conventional logMAR charts. This is desirable when monitoring for change in clinical status owing to neural ocular disease when VA may be measured in clinic with a patient's habitual, rather than best refractive correction. The next chapter aims to test the performance of the MACs in AMD to determine their potential ability to differentiate between neural and optical losses of vision.

7. Visual acuity loss in age-related macular degeneration measured using the Moorfields Acuity Chart

CHAPTER 7

The work discussed within this chapter has been published in a peer reviewed journal;

Shah N., Dakin S. C., Dobinson S., Tufail A., Egan C. A. & Anderson R. S. Visual acuity loss in patients with age-related macular degeneration measured using a novel high-pass letter chart. British Journal of Ophthalmology. 2016;100:1346-52.

7.1 Introduction

AMD affects approximately 8.7% of the population worldwide (Keane et al., 2012) and is the leading cause of blindness in the developed world. Owing to a predicted exponential increase in population ageing, this figure is projected to rise such that by 2040, 288 million people will be affected by AMD (Wong et al., 2014) which has a profound effect on quality of life (Mitchell and Bradley, 2006) and poses a major socioeconomic challenge. In recent years, pharmacological inhibition of vascular endothelial growth factor via intravitreal injection of anti-VEGF ranibizumab (Lucentis) or bevacizumab (Avastin) and more recently aflibercept (Eylea) have offered an opportunity to improve visual outcomes in patients with neovascular AMD (Keane et al., 2012). Although the processes in non-neovascular AMD are not fully understood, a number of different treatments are being investigated which either act to prevent the loss of photoreceptors and retinal pigment epithelial cells (e.g. visual cycle inhibitors) or to suppress inflammation (e.g. complement inhibitors, corticosteroids) (Damico et al., 2012). Patients with early AMD, characterized by the development of small-sized and medium-sized drusen or retinal pigment epithelial abnormalities, and those with intermediate AMD, characterized by medium and large drusen and/or geographic atrophy not involving the centre of the fovea (AREDS, 2001), are often asymptomatic. Central vision loss is most strongly associated with the two forms of late AMD; geographic atrophy involving the fovea and neovascular AMD (Keane et al., 2012, AREDS, 2001, Neelam et al., 2009), although current means of self-monitoring with the Amsler chart have demonstrated a poor sensitivity to central field loss, thought to be due to 'perceptual completion or filling in' (Crossland and Rubin, 2007, Crossland and Bex, 2009). Assuming that any future therapies will be more effective if administered early in the disease process, it becomes all the more important to possess a test that is specifically sensitive to the changes in AMD, while at the same time displaying good repeatability in order to better monitor its progression.

High-contrast VA measurement is the globally accepted assessment of visual function both in routine clinical practice and as an end point for many clinical trials (Csaky et al., 2008) and the ETDRS chart is the current gold standard acuity test for this. As discussed in previous chapters, conventional black-on-white letter charts have the advantage of a large nAFC and thus a low guess rate. Secondly, the SF phase reversals that occur in the presence of optical defocus (Ravikumar et al., 2010, Thorn and Schwartz, 1990) mean that they are excellent tests for detecting refractive error. However, previous studies have demonstrated that VA measured with conventional design letters appears to be relatively poor at indicating visual loss in early and intermediate AMD (Neelam et al., 2009, Qiu and Leat, 2009). A decrease in VA of only two or less letters for patients displaying early age-related maculopathy lesions compared to eyes without such early lesions was reported by Klein et al., (1995). Furthermore, most choroidal neovascular membranes have been shown to originate eccentrically thus having little initial impact on VA (Wong et al., 2008, Liu et al., 2014) even in advanced disease until the fovea is observably affected. The relationship between VA and the severity of retinal changes appears to be inconsistent (Hogg et al., 2003) with reported good VAs in eyes with large lesions and poor VAs in eyes with small lesions (Macular Photocoagulation Study Group, 1994). This appears to be influenced not only by the stage of pathology, whereby the exudative stages generally demonstrate a greater loss of VA compared to those with atrophic disease (Ferris et al., 1984), but also by the location and extent of the lesion (Doris et al., 2001, Hogg et al., 2003) with poorer VA demonstrated with a greater distance from the fovea to healthy retina (Doris et al., 2001). A reduction in VA of more than five letters was part of the criteria used in determining retreatment with ranibizumab during the PrONTO study (Colquitt et al., 2008, Lalwani et al., 2009). Considering TRV values of +/-0.07 to +/-0.19 logMAR even in the absence of any clinical change (Hazel and Elliott, 2002, Vanden Bosch and Wall, 1997, Elliott and Sheridan, 1988, Lovie-Kitchin, 1988, Rosser et al., 2003a, Reeves et al., 1993), with higher variability in patients with AMD having been reported (Patel et al., 2008), these levels of visual deficit cannot reliably be detected and thus cannot be diagnostically meaningful. Indeed, current successful primary endpoints in anti-VEGF clinical trials are taken as the proportion of patients gaining 15 letters (3 logMAR lines) or more or losing fewer than 15 letters (Rofagha et al., 2013, Csaky et al., 2008, Colquitt et al., 2008). If VA test sensitivity and specificity could be improved, smaller changes in VA could be reliably detected and successfully reported.

Several alternative tests of visual function have been considered appropriate for detecting macular disease and for monitoring progression and which certainly better correlate with a patient's perceived visual performance (Latham and Usherwood, 2010, Hazel et al., 2000). Reduced CS has been shown in patients who retain a good level of VA despite the presence of drusen. This loss was particularly observable at the high SF end when compared with normal subjects (Kleiner et al., 1988, Stangos et al., 1995). However, in the target population group, it can be challenging to differentiate CS loss attributable to AMD (Kleiner et al., 1988) from that associated with normal ageing (Owsley, 2011) since CS loss can be similarly affected by both optical and neural problems (Kara et al., 2016, Rubin et al., 1993) and is consequently not disease specific (Latham, 1998). In addition, CS testing has demonstrated poor repeatability (Kara et al., 2016, Thayaparan et al., 2007, Latham, 1998) and has not been adopted into routine clinical practice (Latham, 1998) since correct test administration requires even illumination across the chart which can prove to be difficult to achieve in routine testing environments (Thayaparan et al., 2007). Studies reporting on colour vision deficiencies (Feigl et al., 2004, Midena et al., 1997) and foveal flicker sensitivity (Phipps et al., 2004, Luu et al., 2013) do not appear conclusive, with conflicting results being demonstrated. Reading speed has been examined and a significant reduction in those with visual impairment found (Crossland et al., 2008, Ergun et al., 2003). However, it is difficult to differentiate poor reading speed attributable to neural versus cognitive disease (Elliott et al., 2001) and most studies investigating reading speed in AMD include subjects with already established bilateral disease with central absolute scotomas (Chung et al., 2008, Crossland et al., 2004, Calabrèse et al., 2011, Crossland et al., 2005b). Therefore the role of reading speed in early disease and in aiding diagnosis in terms of sensitivity and specificity is unclear (Rubin, 2013). Other functional tests include microperimetry (Wu et al., 2013, Midena and Pilotto, 2017), but this requires good fixation to provide accurate results which is challenging in this subject group (Bellmann et al., 2004, Crossland et al., 2004), photostress recovery (Wolffsohn et al., 2006) and even functional magnetic resonance imaging (Baseler et al., 2011) but in order for a test to be incorporated into routine clinical settings, they must be straightforward to administer and easily understood by patients.

As discussed earlier in this thesis, previous studies have demonstrated that whilst detection and recognition thresholds for targets such as gratings with the same mean luminance as their background are similar under foveal viewing conditions, under extra-foveal conditions, they are distinctly different (Thibos et al., 1996, Thibos et al., 1987a). Whilst detection of these targets is limited by the filtering effects of the eye's optics, peripheral resolution thresholds are limited by the underlying neural sampling density. Subsequent studies (Anderson and Ennis, 1999, Demirel et al., 2012) and also the previous chapters of this thesis have suggested that this is also true for letters generated in this design (Shah et al., 2012a). The advantage of using letters, which have also displayed a robustness to the effects of optical defocus in normal subjects (Shah et al., 2012a), is that they can be incorporated into a VA chart (the Moorfields Acuity Chart) so that the test is easily administrable and the task already familiar to patients. The previous chapter also revealed better and more consistent TRV values with these charts in uncorrected normal subjects using a variation of termination and scoring techniques (Shah et al., 2014). The primary aim of this study was to investigate how VA using the MAC compared to that measured using conventional letter design charts in both normal subjects and patients with AMD in terms of mean bias and TRV. In addition, the previous study investigated the effect of using a new alternative 10 letter alphabet set (B E H K N P R S X Z) on acuity thresholds, chosen for a number of reasons including their more sampling limited behaviour. Whilst the previous study found no benefit to TRV of these letters over the Sloan set, an intention of this current study was to investigate if any benefit can be gained in using this particular letter set, over the Sloan set, to better reveal visual function deficit in AMD.

7.2 Methods

7.2.1 Visual acuity charts

The eight different letter charts from the previous study, which were graphically designed and digitally produced by a graphic designer at Ulster University, Coleraine, N. Ireland, were used again in this study. These were four charts employing the Sloan letter set (C D H K N O R S V Z), two using a conventional 5 x 5 matrix, black-on-white letter design (C S1 and C S2) and two using a high-pass letter design on a grey background of the same mean luminance as the letters (MAC S1 and MAC S2). The other four charts, again, two in conventional letter design and two using the high-pass letter design, incorporated the different 10-letter alphabet (B E H K N P R S X Z). The luminance of the white components of each chart was 114 cd/m^2 and that of the grey background of the high-pass charts was 50.5 cd/m^2. An image of these eight charts is given in Figure 6.1 in the previous chapter.

7.2.2 Subjects

Ninety normal observers (35 male participants) ranging in age from 20 to 84 years (mean age 46.8 years) were recruited from a primary care optometric practice (Hynes Optometrists, London, UK) and from the staff of Moorfields Eye Hospital, London, UK. Eighty patients (36 males) ranging in age from 59 to 95 (mean 80.6 years) with a range of VAs owing to AMD which was diagnosed in an outpatient retinal therapy clinic at Moorfields Eye Hospital were also recruited. The inclusion criteria for the normal subjects was no significant ocular pathology and best-corrected VA of no worse than 0.20 logMAR. The inclusion criteria for the AMD group was no significant ocular pathology other than AMD which was classified using the Age-Related Eye Disease Study (AREDS) grading system (AREDS, 2001) and the patient was excluded from the study if best-corrected VA fell outside the measurement range of the charts at 4 m. Preliminary tests were conducted in order to determine each participant's eligibility to enrol in the study and also for phenotyping purposes. The first of these was baseline refraction (retinoscopy and subjective) at 4 m using ETDRS Chart R as described in Section 2.2.4, and the mean spherical refractive error was measured to be -0.65D (range

-6.75 to +3.50D) and +0.23D (range -2.25 to +4.25D) for the normal and AMD observers respectively. Ocular examination was performed using slit lamp biomicroscopy and slit lamp binocular indirect ophthalmoscopy. The test eye was randomly assigned for the normal subjects using online random assignment software, described in Section 2.4, with the right eye tested in 44 subjects. Mean VA using chart C S1 for the normal subjects was -0.08 logMAR (range -0.20 to 0.14 logMAR). The test eye in the AMD group was the one with the poorest VA with the right eye tested in 43 subjects. Mean C S1 VA for the AMD subjects was 0.45 logMAR (range -0.02 to 1.20 logMAR). Eight subjects were classified as having early AMD (AREDS category 2) with mean C S1 VA 0.16 logMAR (range -0.02 to 0.28 logMAR), eight with intermediate AMD (AREDS category 3) with mean C S1 VA 0.25 logMAR (range 0.12 to 0.52 logMAR) and the rest (64 subjects) with advanced AMD (AREDS category 4) with mean C S1 VA 0.52 logMAR (range 0.02 to 1.20 logMAR). Ethical approval was obtained for this study from the West London Research Ethics Committee and all procedures adhered to the tenets of the Declaration of Helsinki.

7.2.3 Procedure

Following the preliminary screening tests, best corrected VAs were determined on the test eye of each subject using the eight different acuity charts. These were presented in a randomised order, determined again using an online randomiser described in Section 2.4, in order to control for fatigue and learning effects. The test procedure was similar to that for the previous study with subjects required to identify each letter starting from the top of the chart, with encouragement to guess when unsure as per the recommendations of Smith (2005). Once again, viewing time was kept unrestricted since better acuity thresholds and TRV with increasing exposure duration have been demonstrated by Heinrich et al., (2010). The response for each letter was recorded by the examiner on a pro forma data sheet with a circle drawn around each correctly identified letter and a cross through each incorrect letter. Following the recommendations of the previous study (Shah et al., 2014), VA scores were determined using letter-by-letter scoring with line-based termination of four-or-more errors on a single line and the final VA score was calculated in logMAR using Equation 2.1:

VA_{4m} (logMAR) = (1.2 + 0.1) – C_N*0.02.

7.2.4 Statistical analysis

The Shapiro-Wilk W-test and frequency distribution graphs were used to confirm that the differences in the VA measurements with each chart type (C S1 and C S2, MAC S1 and MAC S2, C N1 and C N2 and lastly MAC N1 and MAC N2) were normally distributed. The methods of Bland and Altman (Bland and Altman, 1986) were used to analyse the mean bias and TRV expressed as the 95% LOA between the two test charts in question and ordinary least squares regression analysis was employed to quantify any potential proportional bias. The GraphPad Prism statistical analysis package (GraphPad Software, Inc., La Jolla, CA) was used for these purposes.

7.3 Results

Mean C S1 acuity measured -0.08 logMAR (range -0.20 to 0.14 logMAR) and mean MAC S1 acuity was found to be 0.06 logMAR (range -0.16 to 0.36 logMAR) for the

90 normal subjects. For the 80 AMD subjects, mean C S1 acuity measured 0.45 logMAR (range -0.02 to 1.20 logMAR) and mean MAC S1 acuity was 0.79 logMAR (range 0.20 to 1.20 logMAR).

The effect of age on the difference in VA measurements using MAC S1 compared to the C S1 chart in the 90 normal subjects was investigated first using linear regression analysis. The results of this are displayed in Figure 7.1, where the circular symbols represent the data for the 90 normal subjects. As per the findings of the previous chapter (Shah et al., 2014), MAC acuity thresholds were found to be 'larger' than those of the conventional black-on-white charts. No statistically significant proportional bias was found to indicate any effect of age on the difference in VA achieved between the two tests ($r^2 = 0.038$, $p = 0.064$, dashed line).

However, as explained in the previous chapter, the VA range on the charts used in the study were extended at the poorer end by an additional two lines compared to the range typically found on ETDRS charts (1.2 logMAR compared to 1.00 logMAR) since the charts were designed with an AMD population in mind. The reason for this was to avoid potential complications created by using charts at different test distances (Dong et al., 2002, Patel et al., 2008) since previous findings suggested that VA thresholds with VOs are generally larger (in letter size terms) compared to those with conventional letters. Furthermore, the last line (-0.30 logMAR) was omitted on the charts, so the last line displayed was -0.20 logMAR. The reason for this was that the printing processes to create the VO letters are complex, particularly at the better VA levels where the black and white lines which make up the stroke width become finer. To achieve accurate representation at such levels would require expertise printing input which was beyond the scope feasible for an initial prototype chart. Consequently, a number of the normal subjects did not reach the full termination criterion of four-or-more errors on a line on the conventional letter charts. To hit a ceiling effect with best corrected VA for normal subjects is not uncommon (Adoh et al., 1992). A separate analysis was therefore conducted on 38 subjects who did satisfy the full termination criterion (green circles on Figure 7.1) in order to determine the potential consequences of this ceiling effect. The mean age of this group was 52.1 years (range 20 to 84 years). Mean C S1 acuity for this group was -0.01 (range -0.12 to 0.14 logMAR) and mean MAC S1 acuity was 0.15 logMAR (range -0.04 to 0.36 logMAR). Again no statistically significant proportional bias was found to indicate any effect of age on the difference in VA achieved between the two tests (r^2 = 0.003, p = 0.740, solid line).

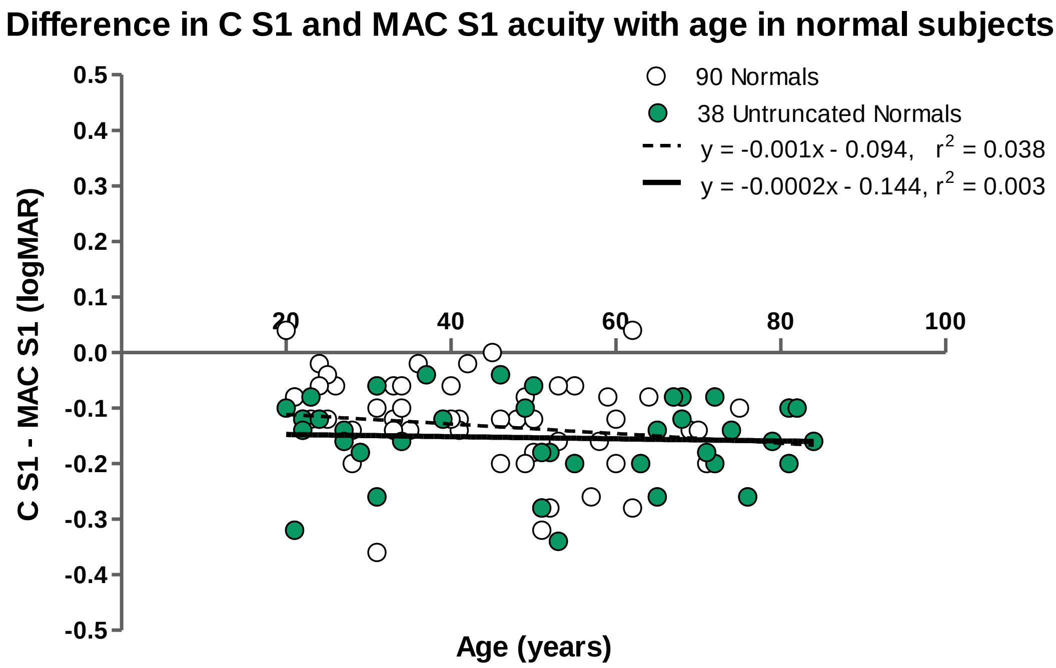


Figure 7.1: Difference in acuity between charts C S1 and MAC S1 against age of subject *(adapted from Shah et al., 2016b)*. Data are shown for all 90 normal subjects (all circular symbols) and for the 38 normal subjects who satisfied the full termination criteria (green filled circles). Regression analysis confirms no significant change in the difference in acuity with increasing age (dashed line for all 90 normal subjects, solid line for the 38 normal subjects who satisfied the full termination criteria).

Since no effect of age on differences in thresholds between the two test types was found when considering all 90 normal subjects or only those who fully satisfied the termination criteria, all ages were included in the subsequent analyses. The Shapiro-Wilk W test was used to test the null hypothesis that the differences in acuity found between charts 1 and 2 of each type in each subject group were sampled from a normal distribution. This was not rejected for C S1 and C S2 in the normal and AMD group (p = 0.086 and p = 0.092 respectively) and for MAC S1 and MAC S2 in the AMD group (p = 0.070). The null hypothesis was rejected for MAC S1 and MAC S2 in the normal group (p = 0.006). A frequency distribution plot to further examine this (Figure 7.2) however, demonstrated no gross deviation from normal and on this basis the use of Bland and Altman summary statistics in terms of mean bias and TRV expressed as the 95% LOA were considered justified.

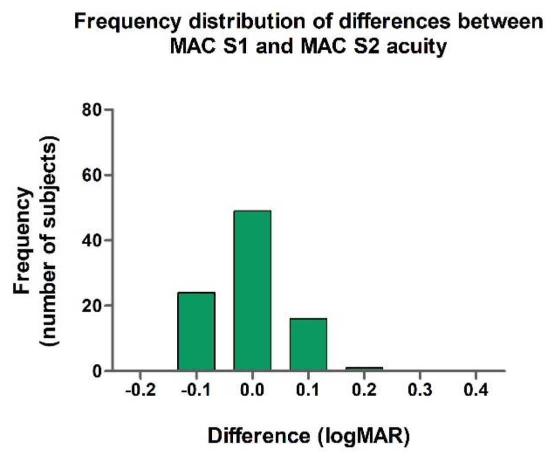


Figure 7.2: Frequency distribution plot of the difference in acuity attained with MAC S1 and MAC S2 charts in the 90 normal subjects. No gross deviation from a normal distribution is demonstrated.

Bland-Altman analysis on all 90 normal subjects revealed a mean bias of -0.01 logMAR between C S1-C S2 measurements with TRV of +/-0.09 logMAR and mean bias of 0.00 logMAR between MAC S1 and MAC S2 with TRV +/-0.11 logMAR. The mean difference between C S1 and MAC S1 was found to be -0.13 logMAR.

However, further comparisons were again only made in the 38 normal subjects who strictly satisfied the termination criteria of four-or-more errors on a line.

The Bland-Altman scatter plots in Figure 7.3 graphically present the distribution of results and TRV (calculated as 1.96 x SD) and mean bias in scores for the (a) C S1 and C S2 charts and (b) MAC S1 and MAC S2 charts for the 38 normal subjects (green data set) and 80 AMD subjects (red data set). These data are also summarised in Table 7.1a. No systematic association between the underlying acuity and the level of agreement between charts 1 and 2 of either the conventional or high-pass design was found. The null hypothesis, that the SD for C S1-C S2 is not significantly different to that of MAC S1-MAC S2, was examined using the F-test. This was not rejected for the normal subjects (two-tailed p = 0.54, $F_{37,37}$ = 1.73) but was rejected for the AMD group (two-tailed p = 0.003, $F_{79,79}$ = 1.45) such that the MAC displayed a lower TRV.

Figure 7.3 (c) and (d) represent the results of a method comparison study giving information on the agreement between the two different chart types (C S1 and MAC S1) in the normal subjects and AMD group respectively. The mean bias between the two test types is -0.15 logMAR (one and a half lines) for the 38 normal subjects with good VA. A much larger mean bias of -0.33 logMAR (approximately three lines) was found in the AMD group. However, from Figure 7.3d, a proportional as well as systematic bias can be inferred such that a greater level of disagreement between the two chart types is evident at the ‘better’ acuity end than the ‘poorer’ acuity end. Ordinary least squares regression analysis confirms this bias (r^2 = 0.133, p = 0.001), such that a difference of -0.45 logMAR (4.5 lines) at the 0.00 logMAR level was found between the charts, with conventional VA measurements again giving ‘better’ acuity results compared to a difference of -0.26 logMAR (approximately two and a half lines) at the 1.00 logMAR acuity level.

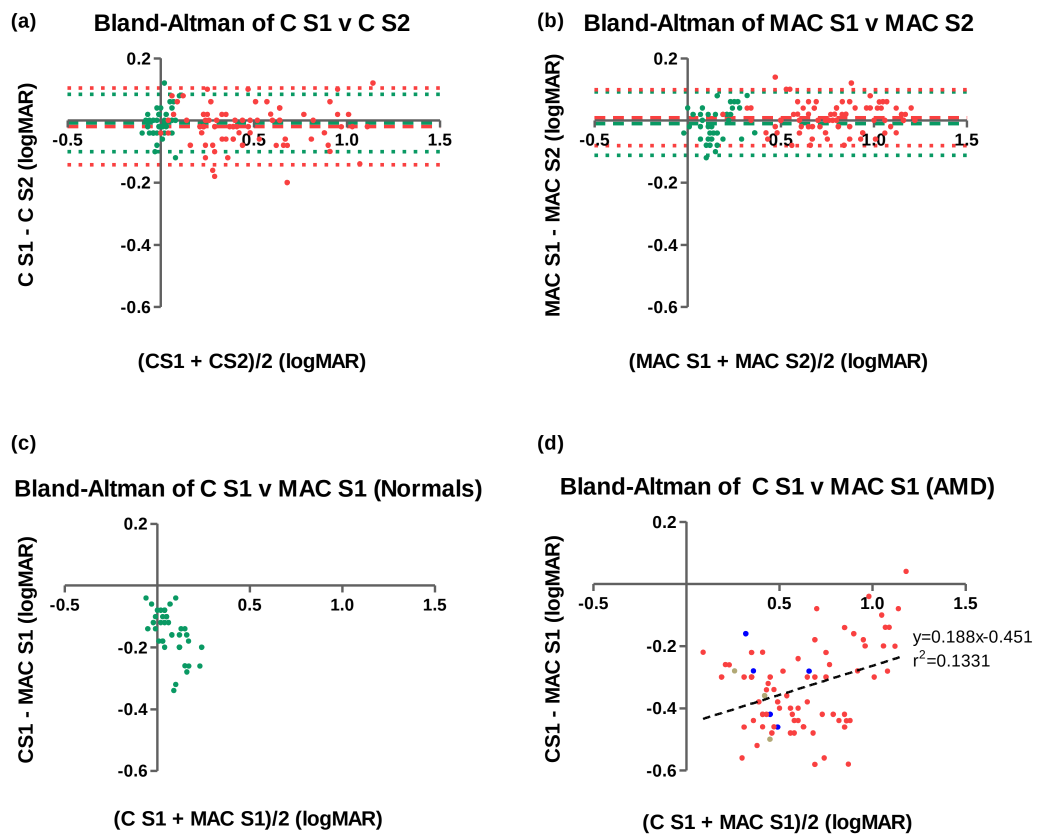


Figure 7.3: Bland-Altman plots for test and retest VA measurements for the normal subjects and AMD patients *(adapted from Shah et al., 2016b).* The mean difference and upper and lower 95% LOA are plotted for (a) C S1 versus C S2 and (b) MAC S1 versus MAC S2 (data for normal subjects in green and the AMD group in red). (c) and (d) display the differences in VA between charts C S1 and MAC S1 in the normal group (n = 38) and AMD group (n = 80) respectively, with the slope fitted using ordinary least squares regression analysis p = 0.001. In (d), the gold, blue and red circles represent subjects with early, intermediate and advanced AMD respectively.

(a)

Normal subjects (n=38)	Mean Difference (SE)	95% CI Mean Difference	Range of Observed Differences	TRV, 95% Limits of Agreement
C S1 – C S2	-0.008 (0.008)	-0.023, 0.008	-0.12, 0.12	±0.092
MAC S1 – MAC S2	-0.010 (0.008)	-0.027, 0.007	-0.12, 0.08	±0.102
AMD subjects (n=80)				
C S1 – C S2	-0.019 (0.007)	-0.033, -0.005	-0.20, 0.12	±0.123
MAC S1 – MAC S2	0.009 (0.005)	-0.001, 0.019	-0.08, 0.14	±0.090

(b)

Normal subjects (n=42)	Mean Difference (SE)	95% CI Mean Difference	Range of Observed Differences	TRV, 95% Limits of Agreement
C N1 – C N2	0.008 (0.007)	-0.007, 0.023	-0.08, 0.14	±0.094
MAC N1 – MAC N2	0.006 (0.008)	-0.010, 0.021	-0.16, 0.08	±0.098
AMD subjects (n=80)				
C N1 – C N2	0.002 (0.006)	-0.011, 0.015	-0.16, 0.14	±0.114
MAC N1 – MAC N2	-0.001 (0.006)	-0.012, 0.011	-0.10, 0.18	±0.099

Table 7.1: Bland-Altman summary statistics for the normal subjects and AMD group *(adapted from Shah et al., 2016b).* The data for the Sloan letter set is shown in (a) where n = 38 for the normal subjects and n = 80 for the AMD group. The data for the New letter set is shown in (b) where n = 42 for the normal subjects and n = 80 for the AMD group. These data are presented graphically in Figure 7.3 and Figure 7.5.

Mean acuity with the New letter set was similar to the Sloan letter set in both the normal subject group and the AMD group. Mean C N1 acuity measured -0.08 logMAR (range -0.20 to 0.10 logMAR) and mean MAC N1 acuity was 0.08 logMAR (range -0.16 to 0.34 logMAR) for the 90 normal subjects. Mean C N1 acuity was 0.47 logMAR (range -0.02 to 1.14 logMAR) and mean MAC N1 acuity was 0.78 logMAR (range 0.26 to 1.20 logMAR) for the AMD group.

As described above, the Shapiro-Wilk W test was used to test the null hypothesis that the differences in acuity with the New letter set, found between charts 1 and 2 of each type in each subject group, were sampled from a normal distribution. This was not rejected for MAC N1 and MAC N2 in the normal group ($p = 0.171$) but was rejected for C N1 and C N2 in the normal and AMD group ($p = 0.038$ and $p = 0.032$ respectively) and for MAC N1 and MAC N2 in the AMD group ($p = 0.003$). Frequency distribution plots to further examine this (shown in Figure 7.4) however, demonstrated no gross deviation from normal and on this basis the use of Bland and Altman summary statistics in terms of mean bias and TRV expressed as the 95% LOA were again considered justified.

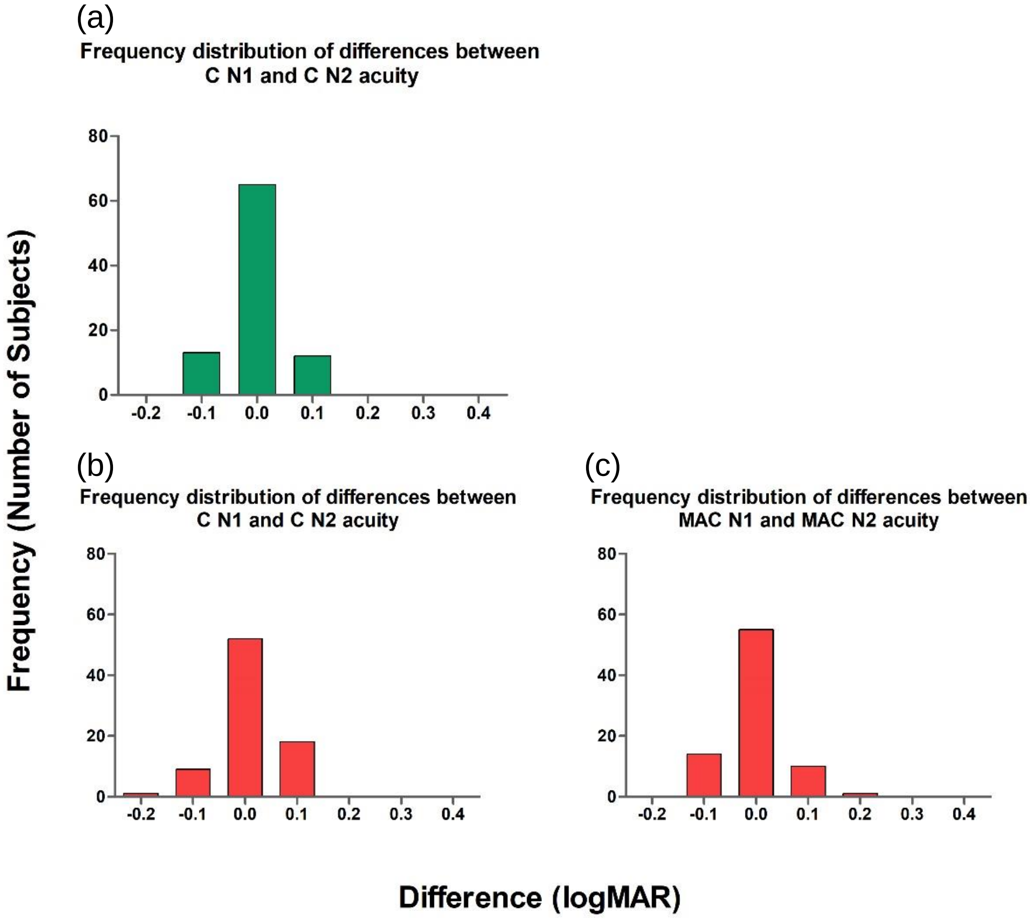


Figure 7.4: Frequency distribution plots of the difference in acuity attained with C N1 and C N2 charts in the (a) 90 normal subjects and (b) 80 AMD patients and (c) MAC N1 and MAC N2 in the 80 AMD patients. No gross deviations from a normal distribution are demonstrated.

Bland-Altman analysis on all 90 normal subjects for the New letter set revealed a mean bias of 0.00 logMAR between C N1 and C N2 measurements with TRV of +/-0.09 logMAR and mean bias 0.01 logMAR between MAC N1 and MAC N2 with TRV +/-0.11 logMAR. The mean difference between C N1 and MAC N1 was found to be -0.16 logMAR. As per the reasoning above, further analysis was made on 42 normal subjects who strictly satisfied the termination criteria of four-or-more errors on a line for the New letter set charts.

The Bland-Altman scatter plots for the (a) C N1 and C N2 charts and (b) MAC N1 and MAC N2 charts for the 42 normal subjects (green data set) and 80 AMD subjects (red data set) are displayed in Figure 7.5 and these data are also summarised in Table 7.1b. The null hypothesis that the SD for C N1-C N2 is not significantly different to that of MAC N1-MAC N2, examined using the F-test, was not rejected for either the normal subjects (two-tailed p = 0.793, $F_{41,41} = 1.68$) nor for the AMD group (two-tailed p = 0.210, $F_{79,79} = 1.45$).

Figure 7.5 gives information on the agreement between the two different chart types (C N1 and MAC N1) in the (c) normal subjects and (d) AMD group. The mean bias between the two test types is -0.17 logMAR for the 42 subjects with good VA with once again, a much larger mean bias of -0.32 logMAR for the AMD group. However, from (d) a proportional as well as systematic bias can be inferred such that again, a greater level of disagreement between the two chart types is evident at the 'better' acuity end than the 'poorer' acuity end. Ordinary least squares regression analysis confirms again, a proportional bias ($r^2 = 0.190$, $p < 0.0001$) such that a difference of -0.44 logMAR at the 0.00 logMAR level was found between the charts, with conventional VA measurements again giving 'better' acuity results compared to a difference of -0.25 logMAR at the 1.00 logMAR acuity level.

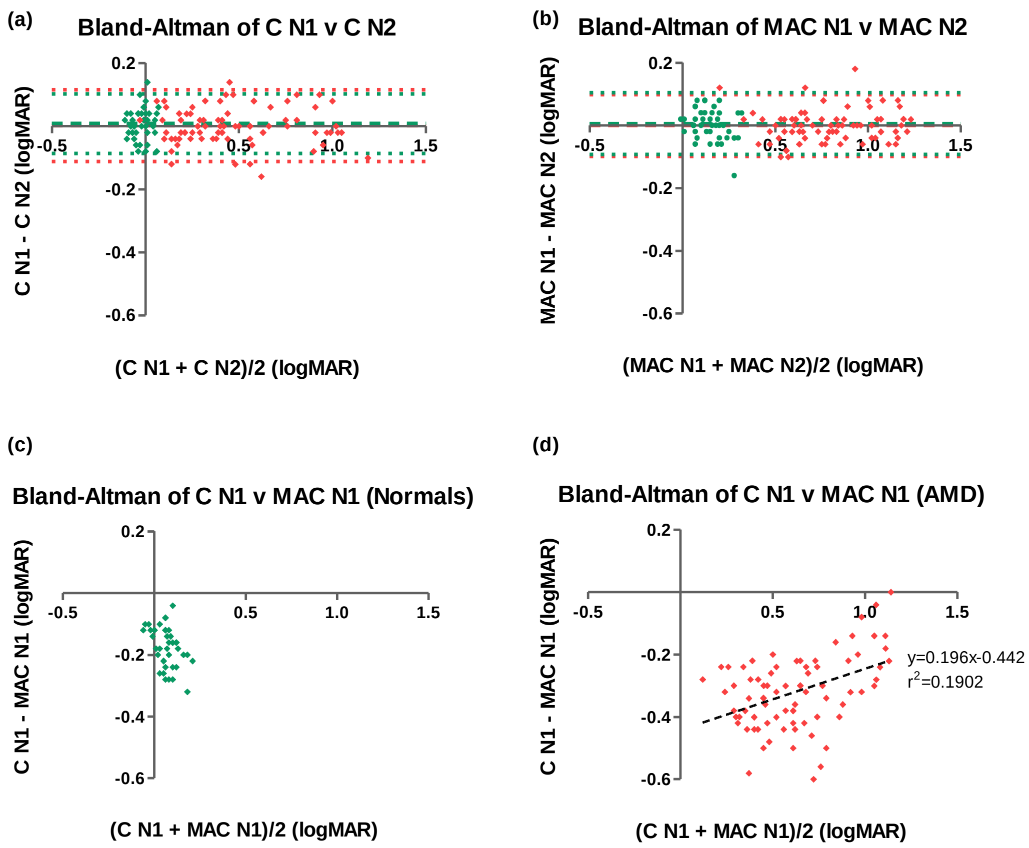


Figure 7.5: Bland-Altman plots for test and retest VA measurements for the New letter set charts in normal subjects and AMD patients. The mean difference and upper and lower 95% LOA are plotted for (a) C N1 versus C N2, (b) MAC N1 versus MAC N2 (data for normal subjects in green and the AMD group in red). (c) and (d) display the differences in acuity between charts C N1 and MAC N1 in the normal group (n = 42) and AMD group (n = 80) respectively with the slope fitted using ordinary least squares regression analysis $p < 0.0001$.

7.4 Discussion

ETDRS VA measurements are now globally accepted as the gold standard for assessing functional changes in patients with AMD and a change in VA is used as the primary outcome measure in clinical trials. The performance of any clinical test designed to detect and monitor for any change in the clinical status is influenced by both its response to the disease 'signal' which it is measuring and the variability of the test (test noise). In reviewing previous literature, it seems apparent that VA measurements can be insensitive to the changes in AMD (Neelam et al., 2009, Qiu and Leat, 2009), even when measured with logMAR charts, with good VA being reported in even advanced disease (Wong et al., 2008, Liu et al., 2014). Previous studies have also demonstrated higher variability in VA measurements in AMD patients compared to normal subjects (Patel et al., 2008, Blackhurst and Maguire, 1989) making it challenging to differentiate true changes in clinical status from test noise. Patel et al., (2008) reported a change (improvement or deterioration) of more than 5 letters in 31% of their subjects with AMD in VA measurements taken one week apart, despite no demonstrable progression of disease. They suggest that TRV in some patients may be explicable by the multiple and inconstant PRLs for fixation which leads to different scores of VA in patients with otherwise no determined change in clinical status (Patel et al., 2008).

A significant proportional bias was found (Figure 7.3 and Figure 7.5) between VA attained with each chart type in the AMD subjects such that, whilst MAC acuity gave larger VA thresholds overall, this difference was larger at the better acuity end than at the poorer end. The difference at the 0.0 logMAR level was found to be approximately 4.5 lines in the AMD patients compared to a difference of 1.5 logMAR lines in normal subjects with similar acuity. Thus, as anticipated, this implies that the MAC which employs high-pass filtered letter targets has a greater response to the disease signal in AMD and thus displays a higher sensitivity in the detection of functional changes which result from neural loss in AMD at levels were ETDRS acuity remains relatively unaffected. The difference between the two charts found here for the normal subjects is similar to the difference of approximately 2 logMAR lines at the good acuity end reported in the previous chapter in normal subjects with uncorrected refractive error (Shah et al., 2014). At the poorer acuity end, the difference in acuity between the charts for the AMD patients was found to be just over 2.5 lines in the present study. This difference which occurs as a result of neurological loss is much larger compared to the difference (approximately half a logMAR line) attributed to optical factors in the previous study at this acuity end (Shah et al., 2014). Thus, this suggests that the MAC is more capable of differentiating between VA losses resulting from neural versus optical deficits in the visual system.

As explained, the signal-to-noise ratio is important to consider, and whilst the MAC appears to have a greater response to the disease signal, this is of no benefit if lost in the TRV. TRV was found to be *clinically* similar overall between chart types measuring +/-0.09 vs. +/-0.10 logMAR for the conventional and MACs respectively for the normal subjects and +/-0.12 vs. +/-0.09 logMAR (which is statistically different) for the AMD patients. The TRV values found for the MAC chart seemed to remain unaffected by the presence of AMD compared to normal subjects (+/-0.09 vs. +/-0.10 logMAR respectively) or uncorrected refractive error (+/-0.10 logMAR) reported in the previous chapter (Shah et al., 2014) with the same test termination rules and scoring techniques. An increase in TRV for the conventional charts with AMD compared to normal subjects (+/-0.12 vs. +/-0.09 logMAR respectively) seen here, is in line with reports from previous studies (Blackhurst and Maguire, 1989). In addition, TRV for the conventional charts was also found to be larger in subjects with uncorrected refractive error (+/-0.14 logMAR) in the previous chapter (Shah et al., 2014) which has also previously been reported in other studies (Rosser et al., 2004).

VA measurements using a New alternative 10 letter set to the Sloan letter set were also investigated in this study. The new letters were chosen, as explained in detail in the previous chapter, to create a set which aimed to reduce TRV by selecting letters with the most similar recognition thresholds. In addition, letters which may potentially be more sensitive to functional deficits in AMD were chosen by selecting those letters which displayed the most sampling limited performance peripherally in high-pass format. The previous study found minimal differences in TRV between the New alternative alphabet set compared to the Sloan set. Similarly, no significant difference to TRV was found in the present study by using the New letter set in both conventional and VO format for both the normal subjects and those patients with AMD. In addition, the differences in VA between the conventional chart and MAC with the New letter set was similar to that using the Sloan set in the normal subjects and in the AMD group (at both the better and poorer ends of acuity). Thus, no additional gain was achieved by using a New letter set over the Sloan letter set in terms of TRV or in the detection of functional deficits owing to AMD. The improvement to both of these seemingly arises from using a high-pass over conventional letter design regardless of the specific letter choices.

Thus, in conclusion, the MAC employing high-pass letters following a logMAR ETDRS chart design and testing and scoring protocols, appears to be more sensitive in detecting functional loss associated with AMD compared to conventional letter design charts. These high-contrast targets have been shown previously to be robust to age related media changes (Chapter 5) which is essential to consider in this target population, and to optical defocus (Chapter 4). This is valuable when considering the potential role of the MACs in differentiating between optical and neural deficits in the visual system. Importantly also, the MAC charts display comparable but also more consistent TRV values than conventional letter charts. As more treatments for AMD come online, this earlier detection within the disease process will become of increasing importance. The next chapter aims to explore further why the MAC should display higher sensitivity to visual loss in AMD.

8. Vanishing Optotype detection and recognition thresholds in age-related macular degeneration

CHAPTER 8

The work discussed within this chapter has been published in a peer reviewed journal;

Shah N., Dakin S. C., Dobinson S., Tufail A., Egan C. A. & Anderson R. S. Visual acuity loss in patients with age-related macular degeneration measured using a novel high-pass letter chart. British Journal of Ophthalmology. 2016;100:1346-52.

8.1 Introduction

The increasing problem that AMD poses on the ever growing ageing population was highlighted in the previous chapter. In non-neovascular or dry AMD, extracellular deposits (drusen) accumulate between the retinal pigment epithelium (RPE) and Bruch's membrane, causing RPE dysfunction and resulting in the degeneration of the photoreceptors (Keane et al., 2012, Ambati and Fowler, 2012), although the process and aetiology behind this is not yet fully understood (Zajac-Pytrus et al., 2015). Eventually the loss of large areas of RPE and outer retina result in geographic atrophy. In neovascular or wet AMD, photoreceptors are destroyed by exudation and blood when sub-choroidal neovascularisation infiltrates through defects in Bruch's membrane and the RPE layer to the sub-retinal space (Zajac-Pytrus et al., 2015). Whilst VA measurement with the ETDRS chart is considered the current gold standard test of visual function, it often appears normal in the early stages of AMD, with VA often only significantly deteriorating in the advanced stages of disease once the fovea becomes involved (Neelam et al., 2009, Qiu and Leat, 2009, Liu et al., 2014). A higher sensitivity to functional visual loss in AMD with the MAC (Shah et al., 2016b) compared to a conventional letter chart was demonstrated in the previous chapter. In addition, a proportional relationship was found such that a larger difference in acuity between conventional black-on-white letter charts and the MAC was found at the better acuity end of the scale than the poorer end, but the reasons for this are yet unknown.

A combination of optical and neural factors limit the ability of the visual system to resolve spatial detail as discussed in Section 1.6. Under foveal viewing conditions in a normal eye, the low-pass, anti-alias filtering effects of the optics of the eye mean that SFs higher than the resolution limit of the retina do not pass through. For high-pass filtered letters presented on a background of same average mean luminance, as the letters approach the recognition threshold when the high SFs can no longer be resolved, the letter also appears to vanish owing to the lack of luminance cue. The neural spatial resolving ability of the retina falls with eccentricity from the fovea at a faster rate than the optical quality of the eye (Green, 1970) and thus detection and recognition thresholds separate for these targets outside foveal vision (Anderson and Ennis, 1999, Demirel et al., 2012, Shah et al., 2012a). However, it may be reasonable to assume a similar process occurs in the development of AMD as the dysfunction of photoreceptors results in a reduced density in the mosaic and/or increased irregularity such that the neural sampling density and not the optics become the ‘weak link’ in the chain of visual processing even under central viewing conditions. If this is the case, then the detection and recognition thresholds for the VO letters should be seen to separate, just as they do for peripheral vision. The aim of this study was to measure these thresholds for the high-pass letters in AMD patients to test this idea experimentally.

8.2 Methods

8.2.1 Subjects

A subset of 19 subjects (12 male) from the previous study were recruited for this part of the study with ETDRS acuity ranging from -0.14 to 0.64 logMAR. Of these, 9 were normal participants, ranging in age from 69 to 81 years (mean age 75.3 years) and 20 were AMD patients, ranging in age from 70 to 90 years (mean age 79.3 years). Within the AMD subgroup, 3 had early AMD (AREDS category 2) with VA ranging from -0.10 to 0.20 logMAR, 2 had intermediate AMD (AREDS category 3) with VA -0.04 and 0.12 logMAR and 15 had advanced AMD (AREDS category 4) with VA ranging from 0.10 to 0.64 logMAR. The test eye was the same eye that was used for the previous study with the right eye tested in 19 of the subjects and the refractive error as determined for 4 m testing was used, with mean spherical refractive error +1.31D (range -2.75D to +4.00D) and +0.33D (range -2.25D to +1.88D) for the normal and AMD subgroups respectively. Ethical approval was obtained for this study from the West London Research Ethics Committee and all procedures adhered to the tenets of the Declaration of Helsinki.

8.2.2 Procedure

VO letters (described in Section 2.1.2) were generated using MATLAB (version 7.6, Mathworks, Inc., Natick, MA, USA) and were presented for a duration of 500 ms using an Apple Macintosh computer (Apple, Inc., Cupertino, CA) on a γ-corrected high resolution (1280 x 1024 pixels) Dell Trinitron P992 CRT monitor (Dell Corp. Ltd, Bracknell, Berkshire, UK). A Bits++ video processor (Cambridge Research Systems, Ltd., Rochester, UK) was used to achieve true14-bit contrast resolution whilst the OpenGL capabilities of the computer's built-in graphics card (ATI Radeon X1600; AMD, Sunnyvale, CA, USA) was used to scale the stimuli. This bilinear interpolation procedure allowed stimuli to be displayed of arbitrary size with sub-pixel resolution whilst retaining their balanced luminance structure. The luminance of the background of the CRT monitor measured 53.6 cd/m^2. All testing was conducted under low room illumination to avoid screen reflections. A viewing distance of 4 m was used and the screen subtended 4.6 x 3.7° and one pixel subtended 0.21 arcmins at this test distance.

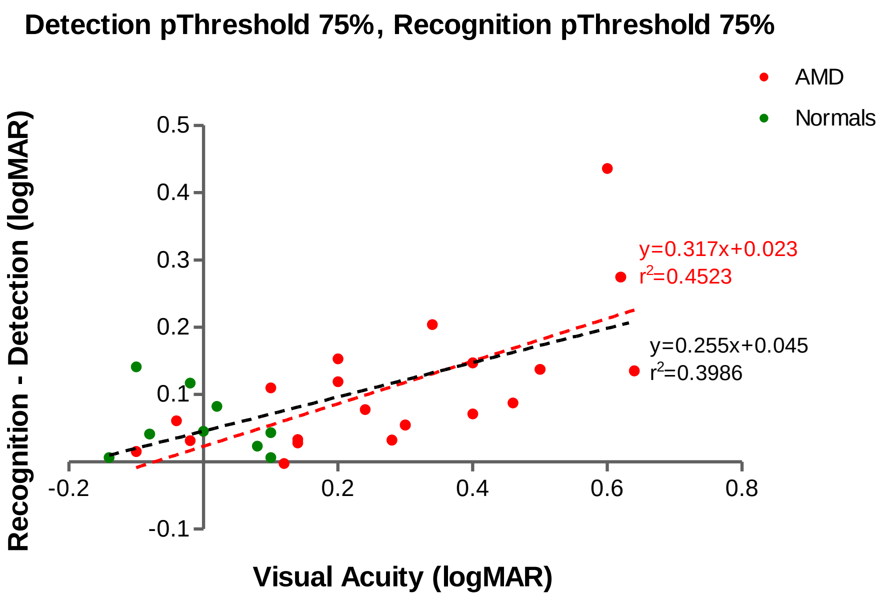
Detection acuity thresholds for the VOs were attained using a temporal 2AFC procedure in which the subject was asked to verbally indicate which of two intervals contained the stimulus high-pass letter. Recognition thresholds were measured using a spatial 10 AFC (the 10 Sloan letters) procedure in which the subject was asked to verbally identify the letter. The subject responses were entered on a keyboard by the examiner. An adaptive forced-choice reversal staircase procedure (QUEST) was employed with the prior density function being limited by the maximum and minimum displayable letter size on the screen. The slope (β) of the psychometric function used was set to 3.5 with the gamma value (guess rate) set to 50% and 10% for the detection and recognition tasks respectively. The pThreshold value was set to 75% correct for the detection and 55% for the recognition task. The subject was not made aware of the 10 letter choices available, although the more astute of the subjects may have been aware of the limited letter set. Since little has previously been reported about a patient's experience with different percent-correct levels and the effect this may have on final thresholds, recognition thresholds were also measured with pThreshold set to 75% since it could be possible that motivation may be better when functioning at a higher percent-correct level where the subject may feel more confident in their responses. Each test run involved 30 trials. Three repeat measures were made for each condition in a random order, determined using an online randomiser (described in Section 2.4) in order to control for fatigue and learning effects, with the final acuity threshold determined by QUEST's built in maximum likelihood estimation procedure of threshold.

8.2.3 Statistical analysis

The final threshold letter size, generated as a percentage with reference to a box size of 512 pixels was converted to a logMAR score for further analysis where, as described in Section 2.1.2, the stroke width for the VO included both the central dark bar and the surrounding white flanks. The GraphPad Prism statistical analysis package (GraphPad Software, Inc., La Jolla, CA) was used to quantify the relationship between the difference in recognition and detection thresholds and VA using ordinary least squares regression analysis.

8.3 Results

The mean of the three repeat threshold measurements for each subject for each condition were calculated. The difference in recognition and detection thresholds in logMAR were calculated and plotted against VA in Figure 8.1, with the top graph showing recognition with pThreshold 75%, and the bottom recognition with pThreshold 55%. No qualitative observable difference in compliance was perceived between both threshold performance levels with full data sets attained for all subjects. The 20 AMD patients are plotted in red with linear regression analysis also plotted in red for this group, and the 9 normal subjects in green, with regression analysis for all subjects plotted as a black dotted line.



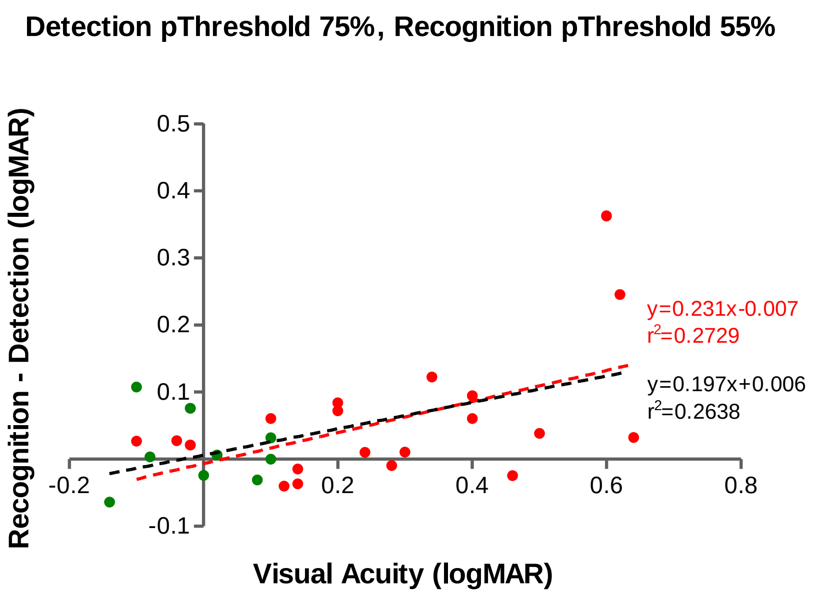


Figure 8.1: Difference in detection and recognition thresholds against VA for 9 normal subjects (green circles) and 20 patients with AMD (red circles) *(adapted from Shah et al., 2016b).* Results are presented with pThreshold converging on 75% for both detection and recognition (top graph) and pThreshold of 75% for detection and 55% for recognition (bottom graph). Linear regression analysis is plotted for the AMD group only (red dotted line) and for all subjects (black dotted line).

At the good acuity end (to the left on the x-axis), little difference in detection and recognition thresholds were found as expected with 'Vanishing' Optotype letters in normal subjects. In the top graph, there does appear to be a slight disparity when recognition thresholds converged on 75% correct with a difference of 0.05 logMAR between detection and recognition when considering all subjects. This is probably due to differences in nAFC between the detection and recognition tasks, since the difference diminishes to 0.01 logMAR when the recognition thresholds converged on 55% correct (bottom graph).

The difference between detection and recognition thresholds increases as acuity worsens owing to AMD with predicted differences in thresholds at VA levels of 1.0 logMAR expected to be 0.3 logMAR with pThreshold for recognition of 75% and 0.2 logMAR with pThreshold for recognition of 55%.

8.4 Discussion

The purpose of this study was to explore why the MAC displays a higher sensitivity to visual loss in AMD than conventional letter acuity as reported in the previous chapter. For foveal vision in a normal healthy eye, visual resolution is limited by the low-pass filtering effects of the optics of the eye, such that SFs higher than the Nyquist limit do not pass through (Campbell and Gubisch, 1966). The sampling theory of visual resolution states that the highest SF which can be faithfully represented by a neural array is one half of the neural sampling frequency for a regular sampling array (as cited in Green, 1970). The quantitative relationship between VA (measured with gratings) and cell density (CD) is described for a square array by Equation 8.1 (Thibos, 1998).

$$VA = 0.5 * \sqrt{D}$$

(Eq. 8.1)

At the level of the retina, the neural sampling limit is set by the photoreceptor array foveally, and thus any degeneration of this resulting from ocular disease should effect VA. It can be calculated from Equation 8.1 that the sampling density must be reduced by approximately 75% for VA to deteriorate by 50%. Whilst the photoreceptor array is more complex and irregular, this relationship is a useful first approximation (Thibos, 1998). Geller et al., (1992) found that grating resolution fell from 100% to 75% threshold performance levels only after more than 88% of the sampling elements were removed. Therefore, it is evident that a disproportionate number of photoreceptors must be lost before VA is noticeably affected. However, this is applicable to periodic sinusoidal gratings where a higher degree of redundancy may be displayed and which may continue to be sampled at a high frequency in undamaged areas amongst the overall degenerating photoreceptor mosaic (Geller et al., 1992). The effects of spatial sampling on letter optotypes, which have more complex Fourier spectra compared to gratings, with lower redundancy, are less well understood (Carkeet et al., 2008). Alexander et al., (1995) reported a marked effect on letter recognition over grating resolution with spatial sampling, achieved by masking pixels in the computer generated stimuli at spatially random positions. However, Seiple et al., (1995) reported a relatively high accuracy in letter identification even with substantial pixel blanking in the stimulus.

Figure 8.2 (a) and (b) simulates the appearance of the two different chart types after they have undergone sampling by a noisy, reduced-density sampling array as might be expected to occur in AMD. Figure 8.2 (c) and (d) simulate a further decline in the sampling density and the appearance of the charts which may be expected following foveal involvement.

It can be appreciated from the simulation that the detection and recognition thresholds for the high-pass letters (Figure 8.2b) begin to separate with under-sampling. The difference in detection and recognition thresholds for the high-pass letters becomes even more apparent with further under-sampling (Figure 8.2d) such that the subject can detect stimulus contrast but can no longer veridically resolve the letter. The results from the investigation of this study (Figure 8.1) supports the prediction of the simulation where it can be seen that little difference was observed between detection and recognition thresholds at the good acuity end, however, as VA deteriorated with AMD, the difference in detection and recognition thresholds increases. This finding suggests that under-sampling as a result of photoreceptor loss may indeed underlie the acuity loss with high-pass filtered letters in AMD and foveal vision begins to behave more like peripheral vision.

The ability to correctly identify a letter is reduced with contour interaction or crowding which is not known to impair detection (Whitney and Levi, 2011). Although associated with being particularly detrimental in amblyopia (Levi, 2008), it has also been demonstrated in visually impaired patients including those with AMD (Pardhan, 1997). Qiu (2009) found that in early AMD, crowding effects are similar to those found in normal subjects and it could be that increased crowding in AMD occurs as a result of the uptake of an eccentric retinal fixation location at which the effects of crowding are greater than at the fovea (Latham and Whitaker, 1996). It may be reasonable to assume that the VO letters induce more within-letter crowding effects owing to their complex structure compared to the conventional letters, however this has not been investigated to date. Nevertheless, there may be some contribution from crowding to the separation of detection and recognition thresholds seen here with worsening VA although this would require further formal investigation. In addition, Cheong et al., (2007) demonstrated longer temporal thresholds for letter recognition in patients with AMD compared to age-matched controls. Thus, this may also further influence the increasing separation of detection and recognition thresholds demonstrated in this study as VA worsens with AMD.

From the simulation in Figure 8.2 (a) and (b), it can also be seen that the MAC chart suffers the effects of under-sampling to a greater extent than the conventional letter chart and a large difference in recognition limits between the two charts can be observed. Previous studies (Evans et al., 2010, Geller et al., 1992) have demonstrated that when sampled by an irregular array, resolution for gratings demonstrate a 'supra-Nyquist performance' whereby performance exceeds that expected from the average sampling limit and is determined by a localised area with the highest sampling density. A similar occurrence may be happening here whereby the range of SFs present in the conventional letters mean that the letters remain resolvable until the majority of the foveal photoreceptors become dysfunctional in AMD. For the high-pass letters however, a much smaller proportion of photoreceptors must become dysfunctional before loss becomes evident. Seiple et al., (1995) suggested that with a reduction in sensory input, higher order processes become important and can aid letter identification such as cognitive completion, whereby familiar stimuli such as letters, can be correctly guessed by filling in missing sensory input. They also reported poorer accuracy with an increasing number of alternative letter choices. A previous study in this thesis reported that high-pass letter recognition thresholds were less affected by the nAFC (Shah et al., 2011a) owing to the greater between-letter similarity and resultant greater uncertainty. This means that the high-pass letters are less likely to be correctly guessed compared to the conventional letters when the visual system uses what information is available following the effects of under-sampling.

With a further decline in the sampling density (Figure 8.2 c and d) simulating foveal involvement, recognition for both charts appears to suffer more equally such that the difference in VA loss between the charts is reduced. This may explain the proportional bias seen in Figure 7.3d in the previous chapter. From this figure, the difference in VA between the MAC and conventional charts in those with poor VAs owing to AMD can be observed to be approximately two and a half lines. This is similar to the differences seen between detection and recognition thresholds in healthy experienced psychophysical observers at 10° eccentricity in the nasal field (Shah et al., 2012a) in Chapter 4 (Table 4.1), although a different methodology was used. Although fixation loci were not charted in this study, it may be that at some point the uptake of a PRL for fixation away from the fovea to an alternative more intact area of the retina in more advanced AMD results in differences between the charts similar to the values found in the peripheral retina.

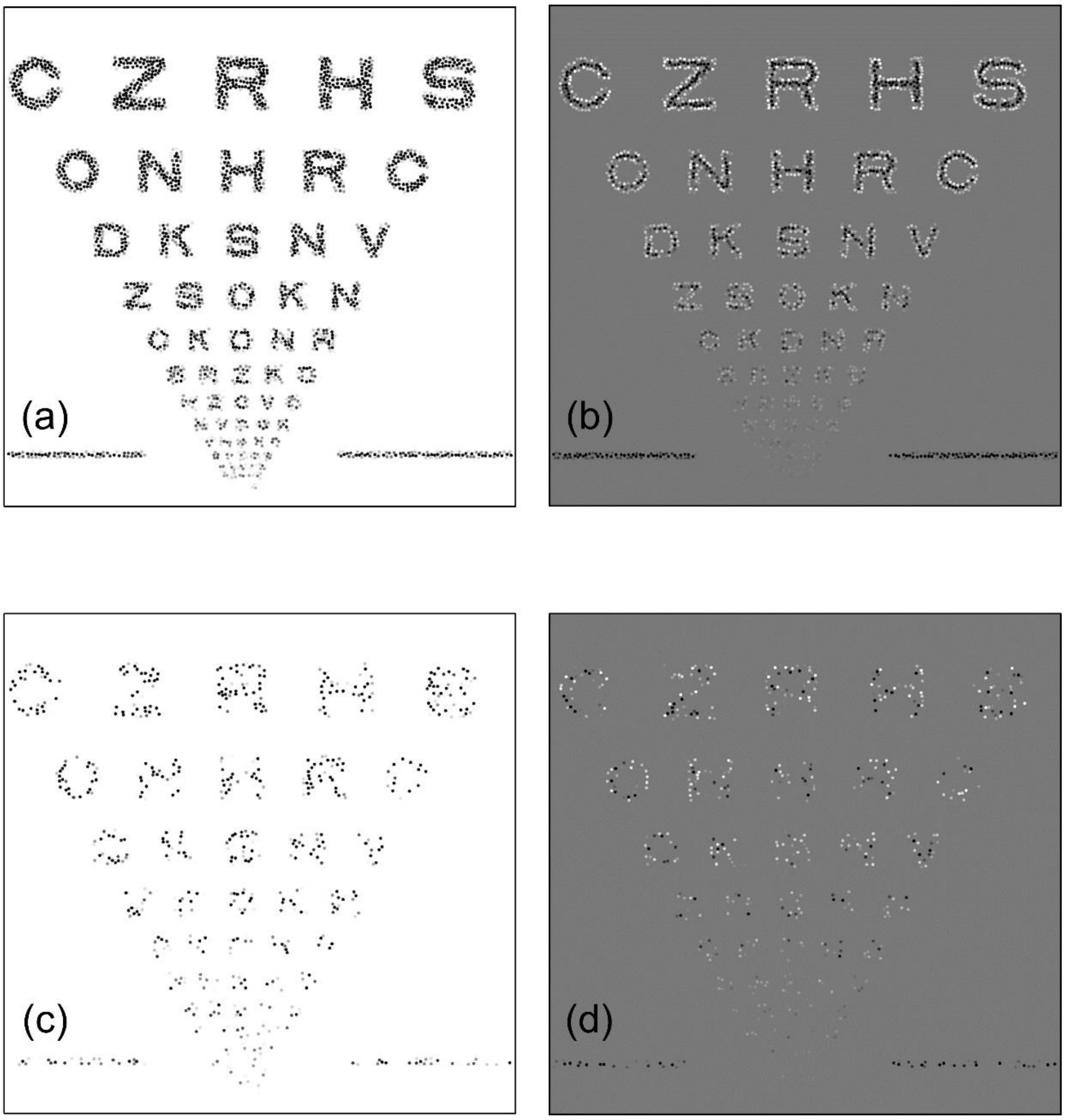


Figure 8.2: Simulation images of the conventional charts and MACs sampled by a noisy, reduced density sampling array *(adapted from Shah et al., 2016b).* The schematic representations of the conventional charts and MACs demonstrates what may be perceived in AMD before (a and b) and after (c and d) the fovea becomes affected.

In conclusion, the findings of this study support the notion that detection and recognition thresholds for the high-pass filtered letters separate as vision worsens with AMD. This is similar to their behaviour in parafoveal vision where recognition thresholds again become limited as a result of under-sampling. Whilst VA measurements with the MAC appear to be more sensitive in detecting functional loss in AMD, the increasing separation of detection and recognition thresholds could potentially be useful in their own right in monitoring or detecting progression of AMD.

9. Thesis discussion and conclusions

CHAPTER 9

Although a seemingly simple test to use in assessing visual function, it is evident that the interpretation of VA measurement results can be challenging and complex. The 'Recognition Pyramid' of Thibos and Bradley (1993) explains the hierarchical system of visual processes that lead to letter recognition in VA testing (Figure 1.9). The issue with current VA tests, employing conventional black-on-white letters in which there is a large difference in mean luminance between the letter and the background, is that it is not clear at which level the pyramid (visual system) fails when a letter is incorrectly identified, without the need for further investigation. Correct letter recognition could be limited by optical inadequacies within the visual system preventing stimulus contrast detection (first layer) or by neural deficits preventing detail resolution. Indeed, the deficiency could even be higher up in the visual processing system whereby cognitive impairment prevents the final stage of letter recognition.

Whilst conventional letters which contain a range of SFs are sensitive to the effects of optical defocus owing to their vulnerability to phase distortions (Thorn and Schwartz, 1990, Nestares et al., 2003, Ravikumar et al., 2010), it has been demonstrated that their response to neural deficits may not be as consistent, demonstrated by the variable relationship between VA and retinal changes seen in AMD for example (Hogg et al., 2003, Macular Photocoagulation Study Group, 1994). In addition, TRV limits the reliability of VA testing in detecting or monitoring change in clinical status resulting from ocular disease or with treatment. In recognising the difficulties that this can pose for clinical decision making, an effort has been made over the years to enhance test designs, such as the move from Snellen to the more uniformly designed logMAR charts, and by refining test procedures by employing single letter scoring for example. However, despite these modifications, TRV still remains a significant issue with even higher TRV reported in the presence of optical defocus (Rosser et al., 2004, Carkeet et al., 2001, Elliott and Sheridan, 1988) and ocular disease (Patel et al., 2008, Laidlaw et al., 2008).

With conventional letter recognition, the visual system has been shown to switch to using the low SF information when the high SF information can no longer be resolved as the letters become smaller (Majaj et al., 2002, Alexander et al., 1994, Chung et al., 2002). A source of test variability is thought to arise from the significant differences in the low SF content of conventional letters (Anderson and Thibos, 1999a, Anderson and Thibos, 1999b, Gervais et al., 1984) resulting in considerable variation in recognition thresholds between letters. Howland et al., (1978) introduced high-pass filtered letters in which the low SFs are effectively filtered out. When presented on a background of the same mean luminance, the detection and recognition thresholds coincide for normal subjects under foveal viewing conditions. Previous work by researchers has confirmed that, under peripheral viewing conditions, these thresholds do indeed separate as resolution acuity is limited by the lower retinal sampling density outside the fovea (Anderson and Ennis, 1999, Demirel et al., 2012). Whilst similar to the behaviour seen with gratings (Thibos, 1998, Thibos et al., 1987a, Wang et al., 1997, Anderson, 1996b), the advantage of using letters lies in the lower guess rates associated with the higher number of alternative choices and the test familiarity. Thus it was hypothesised that any condition affecting the underlying neural sampling pool may also result in a separation of these thresholds, even under central viewing conditions. An additional benefit of the removal of the low SF content in high-pass filtered letters is that between-letter recognition thresholds may become more similar as the SF content becomes less dissimilar, thus potentially reducing TRV.

Whilst high-pass targets are currently used in the Cardiff Acuity test in picture format, no current high-pass letter VA test chart adopting the principles of logMAR testing exists. High-pass ring optotypes are also used in HRP but this test was designed to use detection rather than resolution thresholds as a measure of the underlying neural sampling array which may not be entirely appropriate (Ennis and Johnson, 2002). Thus the purpose of this thesis was to investigate the potential use of high-pass letters in a patient-friendly, easily administrable test of visual function. In order to achieve this, a number of preliminary studies were necessary to further investigate the behaviour of these stimuli more closely.

9.1 Summary of findings

VA tests conventionally incorporate a finite number of alternative choice optotypes. However, it is unclear how many different alternatives an individual patient assumes whilst deciding their response. Indeed, a patient may alter their strategy with repeat measures over time, or even during the test as optotypes much larger than threshold are attempted initially and the test becomes more familiar. The first study in this thesis (Chapter 3) compared the effect that the number of alternative letter choices (2AFC, 4AFC, 6AFC and 26AFC) in conventional and high-pass format had on VA thresholds and TRV. As expected, VA size thresholds were found to be larger for every nAFC condition with the VO letters since the visual system is forced to rely on the only available high SF information for letter identification. VA thresholds in conventional letter format were found to be more significantly affected by the nAFC compared to the VO letters regardless of the assumptions made about the guess rate, and these thresholds were also influenced by what the actual letter choices were. The smallest and largest mean threshold acuity values were both found in the 2AFC conditions, -0.33 and 0.06 logMAR respectively for the conventional letters and 0.01 and 0.17 logMAR respectively for the VO letters for fixed guess rate and pThreshold values. This study also demonstrated lower variability in acuity thresholds determined using the VO rather than conventional letters.

The second study (Chapter 4) served to answer several questions, the first of which was to look at the relative recognition thresholds across the whole alphabet of VO and conventional letters which has not been previously examined. The VO letters were found to display a much lower inter-letter variation in recognition threshold for every letter compared to the conventional letters (inter-letter range of 0.15 versus 0.40 logMAR). The considerable differences in recognition thresholds of the Sloan letters in conventional letter format was also clearly evident (inter-letter range of 0.22 logMAR). Whilst these are arranged in combinations to create lines of average similar difficulty on the ETDRS chart, it was anticipated that a different, more equally legible letter set, could be chosen to potentially reduce TRV which would be investigated in later chapters.

Previous studies have demonstrated the sampling limited nature of grating and VO letter resolution outside the fovea, and their robustness to optical defocus, however these were all examined under low guess rate (2AFC) conditions only. This study confirmed the relative robustness of VO letter compared to conventional letter recognition under 26AFC conditions to optical defocus with foveal viewing (0.28 logMAR/D versus 0.35 logMAR/D change). Additionally, detection and recognition thresholds for the VOs remained separate under peripheral viewing with up to +7D of optical defocus suggesting that under these conditions, recognition of these letters is sampling and not optically limited. Furthermore, this study confirmed that the sampling limited performance of recognition thresholds applies to all 26 individual letters of the alphabet in VO format in peripheral viewing.

Whilst VO letters demonstrated a robust performance to optical defocus, the third study (Chapter 5) considered the effects of wide-angle light scatter simulating lens ageing, since the processes behind image degradation differ between the two. Both conventional letter and VO recognition thresholds were only affected by substantial levels of wide angle light scatter corresponding to levels found on average in those over 90 years old or with significant cataract. This advocates the suitability of VO letters in the development of functional vision tests more specific and sensitive to neural damage. This study also concurs with previous reports that conventional letter acuity measurement does not reflect accurately the symptoms often reported with early lenticular changes.

Using the information from the preliminary laboratory based studies combined with pre-existing knowledge about test chart design, the Moorfields Acuity Chart was created. This chart incorporates the design principles and test procedures of the current gold standard ETDRS logMAR chart but uses high-pass filtered rather than conventional letters. The importance of forced choice techniques in achieving accurate and repeatable VA results is already recognised, and Carkeet (2001) demonstrated the effect of termination and scoring techniques on acuity thresholds with conventional letters. In the fourth study (Chapter 6), the effect of these on MAC and conventional letter chart VA thresholds were compared in normal subjects with a range of VAs owing to uncorrected refractive error. Improved TRV was found using the MAC rather than conventional letter charts (+/-0.10 versus +/-0.14 logMAR with letter-by-letter scoring, line-based termination). Furthermore, MAC acuity thresholds and TRV were found to be less affected by the actual scoring technique and termination criteria used. This would certainly be of benefit when comparing VA results attained by different clinicians, or centres employing different testing criteria. This practical study on real-life subjects also supported the simulation predictions of Carkeet (2001) that a termination criterion of 4-or-more letters wrong on a line should be used with single-letter scoring on the 10AFC ETDRS chart for optimum slope corrected SDs. The recommendations of this study suggest that the same criterion is also applicable to the MACs.

The potential benefits to TRV of using letters with more similar recognition thresholds was also assessed using an alternative 10 letter alphabet set with letters selected from the previous study in Chapter 4. In agreement with the findings of Raasch et al., (1998), no significant improvement was found, suggesting that the biggest improvement to TRV is gained by using letters of a high-pass design.

In addition, a proportional bias between MAC and conventional letter acuity was also demonstrated in this study such that a smaller difference between chart types was found at the poorer acuity end (approximately half a logMAR line) compared to the better acuity end (approximately 2 logMAR lines).

A comparison of VA with the MAC and conventional letter charts in patients with a range of VAs owing to AMD was made in the fifth study (Chapter 7). As predicted, VA measures with the MAC appear to be more sensitive to the functional losses resulting from AMD at a level where conventional acuity remains normal with the difference between charts found to be approximately 4.5 lines in the AMD patients compared to 1.5 lines in normal subjects. Once again, a proportional bias was found in the difference in VA achieved with the two charts but the differences in VA occurring here in patients with neural visual losses, are larger than those demonstrated in subjects with optical visual losses in the previous study. Additionally, whilst TRV for the conventional charts was found to vary somewhat depending on the test population in question, being greater in AMD patients (TRV +/-0.12 logMAR) and in the presence of optical defocus in the previous study (+/-0.14 logMAR) compared to fully refracted normal subjects (+/-0.09 logMAR), TRV for the MAC appeared to remain relatively unaffected. TRV for the MAC was found to be either +/-0.09 or +/-0.10 logMAR for all three subject groups.

The purpose of the sixth study (Chapter 8) was to measure and compare detection and recognition thresholds for the VO letters in a subset group of the AMD patients and normal subjects. These thresholds for the VO letters, whilst similar in normal subjects, were found to separate as VA worsened with AMD, such that central vision behaved like peripheral vision. This confirms that recognition thresholds are neural sampling limited in AMD explaining the greater sensitivity of the MAC over the conventional charts to functional loss in AMD. Simulations of the appearance of the two chart types demonstrates that the MAC suffers to a greater extent after they have undergone sampling by a noisy reduced-density sampling array, but the charts begin to suffer more equally with a further decline in sampling density, which may explain the proportional differences found.

9.2 Thesis conclusions

Acuity measured using conventional letter charts can potentially be affected by both optical and neural losses of vision. However, through this thesis, it can be seen that the extent to which each of these occurs is dependent not only on the eye condition in question but also the stage of pathology. Often, disease is well advanced before this is reflected in the VA score and results in large inconsistencies between patient symptoms and clinical measures of visual function. This limits the diagnostic capability of the test in early stage disease and also its ability to differentiate the origin of the functional visual loss. The work in this thesis has demonstrated that it is possible to attain more sampling limited measures of VA using VO letter recognition thresholds with the potential to better separate optical and neural losses of vision. Acuity thresholds and variability with VO letters were less affected by test design features such as the available number of alternative choices, and testing methods such as scoring and termination criteria compared to conventional letters. Importantly, they are also robust to the effects of different types of optical degradation. The Moorfields Acuity Chart, created through this PhD as the first chart of its kind incorporating high-pass letters and following logMAR ETDRS chart design principles and testing protocols, demonstrated a higher sensitivity to functional loss in AMD when conventional acuity still remained normal, whilst also displaying better TRV results. This ability to earlier detect VA loss in AMD will be invaluable as more treatments become available and the need for better representative outcome measures become essential.

There is a number of conceivable ways in which MAC testing could be incorporated into clinical use. Comparative VA results between the two charts could provide complementary and valuable insights, with the MAC used to detect neural deficits whilst conventional letter charts could be continued to be used to detect refractive or optical deficits. Secondly, the increasing separation of detection and recognition thresholds could potentially be used in the temporal monitoring of disease progression such as with AMD and future electronic versions may better facilitate this. The MACs have undergone CE marking to conform to European Standards and are produced commercially under licence by Peter Allen and Associates at PA Vision Ltd.

9.3 Ongoing/future work

Quantifying the disease signal to noise ratio with the MAC: Following the initial results in Chapter 7, that the MAC may be more sensitive to functional loss in AMD, a larger trial is needed to establish normative thresholds for the MAC in early, moderate and advanced AMD. In addition, it is unknown if the ratio of disease signal (difference in thresholds between normal subjects and patients with AMD) to measurement noise or variability is more favourable with the MAC compared to conventional letter charts at different stages of the AMD process. This will be examined in a study supported by a grant (application number R170019A) from the Moorfields Eye Charity on which I am a co-applicant.

Recognition perimetry using high pass filtered letters in patients with glaucoma: As discussed in the introduction to this thesis (Section 1.7.2), HRP measures detection rather than resolution thresholds of a high-pass ring stimulus as a measure of the underlying RGC density. Studies in this thesis (Shah et al., 2012a) and those of others (Anderson and Ennis, 1999, Demirel et al., 2012) have demonstrated that it is clearly the resolution, not detection thresholds which relate to the underlying neural sample. Ennis and Johnson (2002) confirmed that under the lower target contrast and background conditions that HRP utilises, detection and resolution thresholds once again coincided but due to optical and not sampling limited conditions. Results for a study (grant application number 1973, provided by Fight for Sight and Moorfields Eye Charity supporting this PhD), investigating how detection thresholds of a ring stimulus and recognition thresholds to 10AFC high-pass letters compare to structural measures in 20 normal subjects and 20 patients with glaucoma at 4 locations in the peripheral retina, are currently being analysed.

Exploring the psychophysics of keratoconus using the MAC: Studies in this thesis have demonstrated the robustness of MAC acuity to optical degradations. However, Hess and Carney (1979) discovered an attenuation of high SFs only, in early keratoconus. Corneal stromal thinning and the steepening of the cornea which leads to the development of a conical corneal profile in keratoconus, results in increased higher order aberrations (Alio and Shabayek, 2006, Feizi et al., 2013) and forward light scatter (Jinabhai et al., 2012). A grant (application number ST 14 07 B) from the Moorfields Eye Charity supported a project on which I am a co-applicant, to study the effect of these on MAC measurements in keratoconus. The results of these are currently being analysed.

Electronic MAC: There is a tendency now to moving to computerised VA charts and testing systems. There are numerous advantages of currently available electronic tests over traditional hard copy charts such as those outlined in Section 1.2.5. These include letter randomisation, less degradation of chart materials through handling and the ability to incorporate automated algorithms for better controlled forced choice testing with automatic integration into Electronic Medical Records. There would be advantages to incorporating the MAC into an electronic system beyond those already mentioned. Detection and recognition scores could easily be collected separately, and results collected over time could be graphically presented to aid monitoring of disease progression. Much work would need to be done to ensure that the high-pass letters maintain their balanced luminance structure on commercially available screens and investigations into easy calibration would be required if these were to be adopted into routine clinical practice.

References

ADOH, T. O. & WOODHOUSE, J. M. 1994. The Cardiff acuity test used for measuring visual acuity development in toddlers. *Vision Res,* 34**,** 555-560.

ADOH, T. O., WOODHOUSE, J. M. & ODUWAIYE, K. A. 1992. The Cardiff Test: a new visual acuity test for toddlers and children with intellectual impairment. A preliminary report. *Optom Vis Sci,* 69**,** 427-432.

AKUTSU, H., BEDELL, H. E. & PATEL, S. S. 2000. Recognition thresholds for letters with simulated dioptric blur. *Optom Vis Sci,* 77**,** 524-530.

ALEXANDER, K. R., XIE, W. & DERLACKI, D. J. 1994. Spatial-frequency characteristics of letter identification. *J Opt Soc Am A Opt Image Sci Vis,* 11**,** 2375-2382.

ALEXANDER, K. R., XIE, W. & DERLACKI, D. J. 1997. Visual acuity and contrast sensitivity for individual Sloan letters. *Vision Res,* 37**,** 813-819.

ALEXANDER, K. R., XIE, W., DERLACKI, D. J. & SZLYK, J. P. 1995. Effect of spatial sampling on grating resolution and letter identification. *J Opt Soc Am A Opt Image Sci Vis,* 12**,** 1825-1833.

ALIO, J. L. & SHABAYEK, M. H. 2006. Corneal higher order aberrations: a method to grade keratoconus. *J Refract Surg,* 22**,** 539-545.

AMBATI, J. & FOWLER, B. J. 2012. Mechanisms of age-related macular degeneration. *Neuron,* 75**,** 26-39.

AMESBURY, E. C., GROSSBERG, A. L., HONG, D. M. & MILLER, K. M. 2009. Functional visual outcomes of cataract surgery in patients with 20/20 or better preoperative visual acuity. *J Cataract Refract Surg,* 35**,** 1505-1508.

ANDERSON, R. S. 1996a. Aliasing in peripheral vision for counterphase gratings. *J Opt Soc Am A Opt Image Sci Vis,* 13**,** 2288-2293.

ANDERSON, R. S. 1996b. The selective effect of optical defocus on detection and resolution acuity in peripheral vision. *Curr Eye Res,* 15**,** 351-353.

ANDERSON, R. S. 2006. The psychophysics of glaucoma: improving the structure/function relationship. *Prog Retin Eye Res,* 25**,** 79-97.

ANDERSON, R. S. & ENNIS, F. A. 1999. Foveal and peripheral thresholds for detection and resolution of vanishing optotype tumbling E's. *Vision Res,* 39**,** 4141-4144.

ANDERSON, R. S., REDMOND, T., MCDOWELL, D. R., BRESLIN, K. M. M. & ZLATKOVA, M. B. 2009. The Robustness of Various Forms of Perimetry to Different Levels of Induced Intraocular Stray Light. *Invest Ophthalmol Vis Sci,* 50**,** 4022-4028.

ANDERSON, R. S. & THIBOS, L. N. 1999a. Relationship between acuity for gratings and for tumbling-E letters in peripheral vision. *J Opt Soc Am A Opt Image Sci Vis,* 16**,** 2321-2333.

ANDERSON, R. S. & THIBOS, L. N. 1999b. Sampling limits and critical bandwidth for letter discrimination in peripheral vision. *J Opt Soc Am A Opt Image Sci Vis,* 16**,** 2334-2342.

ANDERSON, R. S. & THIBOS, L. N. 2004. The filtered Fourier difference spectrum predicts psychophysical letter discrimination in the peripheral retina. *Spat Vis,* 17**,** 5-15.

APPLEGATE, R. A. 2000. Limits to vision: can we do better than nature? *J Refract Surg,* 16**,** S547-551.

ARDITI, A. & CAGENELLO, R. 1993. On the statistical reliability of letter-chart visual acuity measurements. *Invest Ophthalmol Vis Sci,* 34**,** 120-129.

AREDS 2001. The Age-Related Eye Disease Study system for classifying age-related macular degeneration from stereoscopic color fundus photographs: the Age-Related Eye Disease Study Report Number 6. *Am J Ophthalmol,* 132**,** 668-681.

ARTIGAS, J. M., FELIPE, A., NAVEA, A., FANDIÑO, A. & ARTIGAS, C. 2012. Spectral Transmission of the Human Crystalline Lens in Adult and Elderly Persons: Color and Total Transmission of Visible LightSpectral Transmittance of the Old Human Crystalline Lens. *Invest Ophthalmol Vis Sci,* 53**,** 4076-4084.

ASSOCIATION OF OPTOMETRISTS. 2015. *Vision Standards* [Online]. London. Available: http://www.aop.org.uk/practitioner-advice/vision-standards/ [Accessed 11/06/2015].

BABIZHAYEV, M. A. 2003. Glare disability and driving safety. *Ophthalmic Res,* 35**,** 19-25.

BAI, H. Q., YAO, L., MENG, X. X., WANG, Y. X. & WANG, D. B. 2011. Visual outcome following intraocular foreign bodies: a retrospective review of 5-year clinical experience. *Eur J Ophthalmol,* 21**,** 98-103.

BAILEY, I. L., BULLIMORE, M. A., RAASCH, T. W. & TAYLOR, H. R. 1991. Clinical grading and the effects of scaling. *Invest Ophthalmol Vis Sci,* 32**,** 422-432.

BAILEY, I. L. & LOVIE-KITCHIN, J. E. 2013. Visual acuity testing. From the laboratory to the clinic. *Vision Res,* 90**,** 2-9.

BAILEY, I. L. & LOVIE, J. E. 1976. New design principles for visual acuity letter charts. *Am J Optom Physiol Opt,* 53**,** 740-745.

BAL, T., COECKELBERGH, T., VAN LOOVEREN, J., ROZEMA, J. J. & TASSIGNON, M. J. 2011. Influence of cataract morphology on straylight and contrast sensitivity and its relevance to fitness to drive. *Ophthalmologica,* 225**,** 105-111.

BANISTER, H. 1927. BLOCK CAPITAL LETTERS AS TESTS OF VISUAL ACUITY. *The Br J Ophthalmol,* 11**,** 49-62.

BASELER, H. A., GOUWS, A., CROSSLAND, M. D., LEUNG, C., TUFAIL, A., RUBIN, G. S. & MORLAND, A. B. 2011. Objective visual assessment of antiangiogenic treatment for wet age-related macular degeneration. *Optom Vis Sci,* 88**,** 1255-1261.

BECK, R. W., MAGUIRE, M. G., BRESSLER, N. M., GLASSMAN, A. R., LINDBLAD, A. S. & FERRIS, F. L. 2007. Visual acuity as an outcome measure in clinical trials of retinal diseases. *Ophthalmology,* 114**,** 1804-1809.

BECK, R. W., MOKE, P. S., TURPIN, A. H., FERRIS, F. L., 3RD, SANGIOVANNI, J. P., JOHNSON, C. A., BIRCH, E. E., CHANDLER, D. L., COX, T. A., BLAIR, R. C. & KRAKER, R. T. 2003. A computerized method of visual acuity testing: adaptation of the early treatment of diabetic retinopathy study testing protocol. *Am J Ophthalmol,* 135**,** 194-205.

BECKER, R., TEICHLER, G. & GRAF, M. 2007. Reproducibility of visual acuity assessment in normal and low visual acuity. *Strabismus,* 15**,** 3-6.

BECKER, R., TEICHLER, G. & GRAF, M. 2011. [Comparison of visual acuity measured using Landolt-C and ETDRS charts in healthy subjects and patients with various eye diseases]. *Klin Monbl Augenheilkd,* 228**,** 864-867.

BECKMANN, P. J. & LEGGE, G. E. 2002. Preneural limitations on letter identification in central and peripheral vision. *J Opt Soc Am A Opt Image Sci Vis,* 19**,** 2349-2362.

BELLMANN, C., FEELY, M., CROSSLAND, M. D., KABANAROU, S. A. & RUBIN, G. S. 2004. Fixation stability using central and pericentral fixation targets in patients with age-related macular degeneration. *Ophthalmology,* 111**,** 2265-2270.

BENNETT, A. G. 1965. Ophthalmic test types. A review of previous work and discussions on some controversial questions. *Br J Physiol Opt,* 22**,** 238-271.

BENNETT, A. G. 1986. An historical review of optometric principles and techniques. *Ophthalmic Physiol Opt,* 6**,** 3-21.

BERGIN, C., REDMOND, T., NATHWANI, N., VERDON-ROE, G. M., CRABB, D. P., ANDERSON, R. S. & GARWAY-HEATH, D. F. 2011. The Effect of Induced Intraocular Straylight on Perimetric Tests. *Invest Ophthalmol Vis Sci,* 52**,** 3676-3682.

BIRT, C. M., SHIN, D. H., MCCARTY, B., KIM, C., LEE, D. T. & CHUNG, H. S. 1998. Comparison between high-pass resolution perimetry and differential light sensitivity perimetry in patients with glaucoma. *J Glaucoma,* 7**,** 111-116.

BLACKHURST, D. W. & MAGUIRE, M. G. 1989. Reproducibility of refraction and visual acuity measurement under a standard protocol. The Macular Photocoagulation Study Group. *Retina,* 9**,** 163-169.

BLAND, J. M. 2004. *How can I decide the sample size for a study of agreement between two methods of measurement?* [Online]. Available: http://www-users.york.ac.uk/~mb55/meas/sizemeth.htm [Accessed 26/04/2012].

BLAND, J. M. & ALTMAN, D. G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet,* 1**,** 307-310.

BLAND, J. M. & ALTMAN, D. G. 1999. Measuring agreement in method comparison studies. *Stat Methods Med Res,* 8**,** 135-160.

BOHIGIAN, G. M. 2008. An ancient eye test--using the stars. *Surv Ophthalmol,* 53**,** 536-539.

BOKINNI, Y., SHAH, N., MAGUIRE, O. & LAIDLAW, D. A. H. 2015. Performance of a computerised visual acuity measurement device in subjects with age-related macular degeneration: comparison with gold standard ETDRS chart measurements. *Eye,* 29**,** 1085-1091.

BONDARKO, V. & DANILOVA, M. Crowding effect: different surroundings and different sizes of test stimuli. Perception, 1995. PION LTD 207 BRONDESBURY PARK, LONDON NW2 5JN, ENGLAND, 124-124.

BONDARKO, V. M. & DANILOVA, M. V. 1997. What spatial frequency do we use to detect the orientation of a Landolt C? *Vision Res,* 37**,** 2153-2156.

BRITISH STANDARDS INSTITUTE 2003. BS 4274-1:2003 Visual acuity test types. Test charts for clinical determinations of distance visual acuity. *British Standards Institute*.

BROWN, B. & LOVIE-KITCHIN, J. 1993. Repeated visual acuity measurement: establishing the patient's own criterion for change. *Optom Vis Sci,* 70**,** 45-53.

CALABRÈSE, A., BERNARD, J.-B., HOFFART, L., FAURE, G., BAROUCH, F., CONRATH, J. & CASTET, E. 2011. Wet versus Dry Age-Related Macular Degeneration in Patients with Central Field Loss: Different Effects on Maximum Reading Speed. *Invest Ophthalmol Vis Sci,* 52**,** 2417-2424.

CAMPBELL, F. W. & GREEN, D. G. 1965. Optical and retinal factors affecting visual resolution. *J Physiol,* 181**,** 576-593.

CAMPBELL, F. W. & GUBISCH, R. W. 1966. Optical quality of the human eye. *J Physiol,* 186**,** 558-578.

CAMPBELL, F. W. & ROBSON, J. G. 1968. Application of Fourier analysis to the visibility of gratings. *J Physiol,* 197**,** 551-566.

CAPPE, C., CLARKE, A., MOHR, C. & HERZOG, M. H. 2014. Is there a common factor for vision? *J Vis,* 14**,** 1-11.

CARKEET, A. 2001. Modeling logMAR visual acuity scores: effects of termination rules and alternative forced-choice options. *Optom Vis Sci,* 78**,** 529-538.

CARKEET, A., GERASIMOU, D. F., PARSONSON, L. R., BIFFIN, K. L. & FREDERICKSEN, B. J. 2008. Thresholds for sampled sloan letters are smaller than sample spacing. *Optom Vis Sci,* 85**,** 1142-1150.

CARKEET, A., LEE, L., KERR, J. R. & KEUNG, M. M. 2001. The slope of the psychometric function for Bailey-Lovie letter charts: defocus effects and implications for modeling letter-by-letter scores. *Optom Vis Sci,* 78**,** 113-121.

CERVINO, A., MONTES-MICO, R. & HOSKING, S. L. 2008. Performance of the compensation comparison method for retinal straylight measurement: effect of patient's age on repeatability. *Br J Ophthalmol,* 92**,** 788-791.

CHARMAN, W. N. 2006. Spatial frequency content of the Cardiff and related acuity tests. *Ophthalmic Physiol Opt,* 26**,** 5-12.

CHAUHAN, B. C. 2000. The value of high-pass resolution perimetry in glaucoma. *Curr Opin Ophthalmol,* 11**,** 85-89.

CHAUHAN, B. C. & HOUSE, P. H. 1991. Intratest variability in conventional and high-pass resolution perimetry. *Ophthalmology,* 98**,** 79-83.

CHAUHAN, B. C., HOUSE, P. H., MCCORMICK, T. A. & LEBLANC, R. P. 1999. Comparison of conventional and high-pass resolution perimetry in a prospective study of patients with glaucoma and healthy controls. *Arch Ophthalmol,* 117**,** 24-33.

CHENG, Y., SHI, X., CAO, X. G., LI, X. X. & BAO, Y. Z. 2013. Correlation between contrast sensitivity and the lens opacities classification system III in age-related nuclear and cortical cataracts. *Chin Med J (Engl),* 126**,** 1430-1435.

CHEONG, A. M. Y., LEGGE, G. E., LAWRENCE, M. G., CHEUNG, S.-H. & RUFF, M. A. 2007. Relationship between slow visual processing and reading speed in people with macular degeneration. *Vision Res,* 47**,** 2943-2955.

CHUA, B. E., MITCHELL, P. & CUMMING, R. G. 2004. Effects of cataract type and location on visual function: the Blue Mountains Eye Study. *Eye (Lond),* 18**,** 765-772.

CHUI, T. Y., YAP, M. K., CHAN, H. H. & THIBOS, L. N. 2005. Retinal stretching limits peripheral visual acuity in myopia. *Vision Res,* 45**,** 593-605.

CHUNG, S. T. L., JARVIS, S. H., WOO, S. Y., HANSON, K. & JOSE, R. T. 2008. Reading Speed Does Not Benefit from Increased Line Spacing in AMD Patients. *Optom Vis,* 85**,** 827-833.

CHUNG, S. T. L., LEGGE, G. E. & TJAN, B. S. 2002. Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision Res,* 42**,** 2137-2152.

COLENBRANDER, A. 2001. Measuring vision and vision loss. *In: Duane TD, Tasman W, Edward A. Duane's Clinical Ophthalmology Vol 5.* Philadelphia: Lippincott, Williams & Wilkins.

COLLEGE OF OPTOMETRISTS. 2015. *Test charts* [Online]. London. Available: http://www.college-optometrists.org/en/college/museyeum/online_exhibitions/optical_instruments/charts.cfm [Accessed 22/03/2015].

COLQUITT, J. L., JONES, J., TAN, S. C., TAKEDA, A., CLEGG, A. J. & PRICE, A. 2008. Ranibizumab and pegaptanib for the treatment of age-related macular degeneration: a systematic review and economic evaluation. *Health Technol Assess,* 12**,** iii-iv, ix-201.

COTTER, S. A., CHU, R. H., CHANDLER, D. L., BECK, R. W., HOLMES, J. M., RICE, M. L., HERTLE, R. W., BIRCH, E. E. & MOKE, P. S. 2003. Reliability of the electronic early treatment diabetic retinopathy study testing protocol in children 7 to <13 years old. *Am J Ophthalmol,* 136**,** 655-661.

COUSENS, S. N., ROSSER, D. A., MURDOCH, I. E. & LAIDLAW, D. A. 2004. A simple model to predict the sensitivity to change of visual acuity measurements. *Optom Vis Sci,* 81**,** 673-677.

CROSSLAND, M. & RUBIN, G. 2007. The Amsler chart: absence of evidence is not evidence of absence. *Br J Ophthalmol,* 91**,** 391-393.

CROSSLAND, M. D. & BEX, P. J. 2009. Spatial alignment over retinal scotomas. *Invest Ophthalmol Vis Sci,* 50**,** 1464-1469.

CROSSLAND, M. D., CRABB, D. P. & RUBIN, G. S. 2011a. Task-specific fixation behavior in macular disease. *Invest Ophthalmol Vis Sci,* 52**,** 411-416.

CROSSLAND, M. D., CULHAM, L. E., KABANAROU, S. A. & RUBIN, G. S. 2005a. Preferred retinal locus development in patients with macular disease. *Ophthalmology,* 112**,** 1579-1585.

CROSSLAND, M. D., CULHAM, L. E. & RUBIN, G. S. 2004. Fixation stability and reading speed in patients with newly developed macular disease. *Ophthalmic Physiol Opt,* 24**,** 327-333.

CROSSLAND, M. D., CULHAM, L. E. & RUBIN, G. S. 2005b. Predicting reading fluency in patients with macular disease. *Optom Vis Sci,* 82**,** 11-17.

CROSSLAND, M. D., ENGEL, S. A. & LEGGE, G. E. 2011b. The preferred retinal locus in macular disease: toward a consensus definition. *Retina,* 31**,** 2109-2114.

CROSSLAND, M. D., LEGGE, G. E. & DAKIN, S. C. 2008. The development of an automated sentence generator for the assessment of reading speed. *Behav Brain Funct,* 4**,** 14.

CSAKY, K. G., RICHMAN, E. A. & FERRIS, I. I. I. F. L. 2008. Report from the NEI/FDA Ophthalmic Clinical Trial Design and Endpoints Symposium *. *Invest Ophthalmol Vis Sci,* 49**,** 479-489.

CURRIE, Z., BHAN, A. & PEPPER, I. 2000. Reliability of Snellen charts for testing visual acuity for driving: prospective study and postal questionnaire. *Bmj,* 321**,** 990-992.

DAIEN, V., NAVARRE, S., FESLER, P., VERGELY, L., VILLAIN, M. & SCHNEIDER, C. 2012. Visual acuity outcome and predictive factors after bevacizumab for central retinal vein occlusion. *Eur J Ophthalmol,* 22**,** 1013-1018.

DAMICO, F. M., GASPARIN, F., SCOLARI, M. R., PEDRAL, L. S. & TAKAHASHI, B. S. 2012. New approaches and potential treatments for dry age-related macular degeneration. *Arq Bras Oftalmol,* 75**,** 71-76.

DE WAARD, P. W., JK, I. J., VAN DEN BERG, T. J. & DE JONG, P. T. 1992. Intraocular light scattering in age-related cataracts. *Invest Ophthalmol Vis Sci,* 33**,** 618-625.

DE WIT, G. C., FRANSSEN, L., COPPENS, J. E. & VAN DEN BERG, T. J. 2006. Simulating the straylight effects of cataracts. *J Cataract Refract Surg,* 32**,** 294-300.

DEMIREL, S., ANDERSON, R. S., DAKIN, S. C. & THIBOS, L. N. 2012. Detection and resolution of vanishing optotype letters in central and peripheral vision. *Vision Res,* 59**,** 9-16.

DENNETT, W. S. 1885. Test Type. *Trans Am Ophthalmol Soc,* 4**,** 133-139.

DEPARTMENT OF HEALTH. 2013. *Registering vision impairment as a disability* [Online]. London. Available: https://www.gov.uk/government/publications/guidance-published-on-registering-a-vision-impairment-as-a-disability [Accessed 11/06/2013].

DONG, L., HAWKINS, B. S. & MARSH, M. J. 2002. Consistency between visual acuity scores obtained at different test distances: Theory vs observations in multiple studies. *Arch Ophthalmol,* 120**,** 1523-1533.

DORIS, N., HART, P. M., CHAKRAVARTHY, U., MCCLELAND, J., STEVENSON, M., HUDSON, C. & JACKSON, J. 2001. Relation between macular morphology and visual function in patients with choroidal neovascularisation of age related macular degeneration. *Br J Ophthalmol,* 85**,** 184-188.

EHRMANN, K., FEDTKE, C. & RADIC, A. 2009. Assessment of computer generated vision charts. *Cont Lens Anterior Eye,* 32**,** 133-140.

ELLIOTT, D. B. & BULLIMORE, M. A. 1993. Assessing the reliability, discriminative ability, and validity of disability glare tests. *Invest Ophthalmol Vis Sci,* 34, 108-119.

ELLIOTT, D. B., BULLIMORE, M. A., PATLA, A. E. & WHITAKER, D. 1996. Effect of a cataract simulation on clinical and real world vision. *Br J Ophthalmol,* 80, 799-804.

ELLIOTT, D. B., HURST, M. A. & WEATHERILL, J. 1990. Comparing clinical tests of visual function in cataract with the patient's perceived visual disability. *Eye (Lond),* 4 (Pt 5), 712-717.

ELLIOTT, D. B., HURST, M. A. & WEATHERILL, J. 1991. Comparing clinical tests of visual loss in cataract patients using a quantification of forward light scatter. *Eye,* 5, 601-606.

ELLIOTT, D. B., PATEL, B. & WHITAKER, D. 2001. Development of a Reading Speed Test for Potential-Vision Measurements. *Invest Ophthalmol Vis Sci,* 42, 1945-1949.

ELLIOTT, D. B. & SHERIDAN, M. 1988. The use of accurate visual acuity measurements in clinical anti-cataract formulation trials. *Ophthalmic Physiol Opt,* 8, 397-401.

ELLIOTT, D. B., YANG, K. C. & WHITAKER, D. 1995. Visual acuity changes throughout adulthood in normal, healthy eyes: seeing beyond 6/6. *Optom Vis Sci,* 72**,** 186-191.

ELLIOTT, M. C. & FIRTH, A. Y. 2009. The logMAR Kay picture test and the logMAR acuity test: a comparative study. *Eye (Lond),* 23**,** 85-88.

ELZE, T. & TANNER, T. G. 2012. Temporal properties of liquid crystal displays: implications for vision science experiments. *PLoS One,* 7**,** e44048.

ENNIS, F. A. & JOHNSON, C. A. 2002. Are high-pass resolution perimetry thresholds sampling limited or optically limited? *Optom Vis Sci,* 79**,** 506-511.

ERGUN, E., MAAR, N., RADNER, W., BARBAZETTO, I., SCHMIDT-ERFURTH, U. & STUR, M. 2003. Scotoma size and reading speed in patients with subfoveal occult choroidal neovascularization in age-related macular degeneration. *Ophthalmology,* 110**,** 65-69.

EVANS, D. W., WANG, Y., HAGGERTY, K. M. & THIBOS, L. N. 2010. Effect of sampling array irregularity and window size on the discrimination of sampled gratings. *Vision Res,* 50**,** 20-30.

FALKENSTEIN, I. A., COCHRAN, D. E., AZEN, S. P., DUSTIN, L., TAMMEWAR, A. M., KOZAK, I. & FREEMAN, W. R. 2008. Comparison of visual acuity in macular degeneration patients measured with snellen and early treatment diabetic retinopathy study charts. *Ophthalmology,* 115**,** 319-323.

FARIZA, E., KRONHEIM, J., MEDINA, A. & KATSUMI, O. 1990. Testing visual acuity of children using vanishing optotypes. *Jpn J Ophthalmol,* 34**,** 314-319.

FEIGL, B., BROWN, B., LOVIE-KITCHIN, J. & SWANN, P. 2004. Cone-mediated multifocal electroretinogram in early age-related maculopathy and its relationships with subjective macular function tests. *Curr Eye Res,* 29**,** 327-336.

FEIZI, S., EINOLLAHI, B., RAMINKHOO, A. & SALEHIRAD, S. 2013. Correlation between Corneal Topographic Indices and Higher-Order Aberrations in Keratoconus. *J Ophthalmic Vis Res,* 8**,** 113-118.

FERRIS, F. L., 3RD, FREIDLIN, V., KASSOFF, A., GREEN, S. B. & MILTON, R. C. 1993. Relative letter and position difficulty on visual acuity charts from the Early Treatment Diabetic Retinopathy Study. *Am J Ophthalmol,* 116**,** 735-740.

FERRIS, F. L., 3RD, KASSOFF, A., BRESNICK, G. H. & BAILEY, I. 1982. New visual acuity charts for clinical research. *Am J Ophthalmol,* 94**,** 91-96.

FERRIS, F. L., III, FINE, S. L. & HYMAN, L. 1984. Age-related macular degeneration and blindness due to neovascular maculopathy. *Arch Ophthalmol,* 102**,** 1640-1642.

FRANSSEN, L., COPPENS, J. E. & VAN DEN BERG, T. J. 2006. Compensation comparison method for assessment of retinal straylight. *Invest Ophthalmol Vis Sci,* 47**,** 768-776.

FRIEDENWALD, J. S. 1924. A New Astigmatic Chart. *Am J Ophthalmol,* 7**,** 8-15.

FRISEN, L. 1986. Vanishing optotypes. New type of acuity test letters. *Arch Ophthalmol,* 104**,** 1194-1198.

FRISEN, L. 1987. High-pass resolution targets in peripheral vision. *Ophthalmology,* 94**,** 1104-1108.

FRISEN, L. 1988. Acuity perimetry: estimation of neural channels. *Int Ophthalmol,* 12**,** 169-174.

FRISEN, L. 1991. High-pass resolution perimetry and age-related loss of visual pathway neurons. *Acta Ophthalmol (Copenh),* 69**,** 511-515.

FRISEN, L. 1993. High-pass resolution perimetry. A clinical review. *Doc Ophthalmol,* 83**,** 1-25.

GALLIOT, F., PATEL, S. R. & COCHENER, B. 2016. Objective Scatter Index: Working Toward a New Quantification of Cataract? *J Refract Surg,* 32**,** 96-102.

GEER, I. & WESTALL, C. A. 1996. A comparison of tests to determine acuity deficits in children with amblyopia. *Ophthalmic Physiol Opt,* 16**,** 367-374.

GELLER, A. M., SIEVING, P. A. & GREEN, D. G. 1992. Effect on grating identification of sampling with degenerate arrays. *J Opt Soc Am A,* 9**,** 472-477.

GERVAIS, M. J., HARVEY, L. O., JR. & ROBERTS, J. O. 1984. Identification confusions among letters of the alphabet. *J Exp Psychol Hum Percept Perform,* 10**,** 655-666.

GHASEMI, A. & ZAHEDIASL, S. 2012. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *Int J Endocrinol Metab,* 10**,** 486-489.

GHODRATI, M., MORRIS, A. P. & PRICE, N. S. 2015. The (un)suitability of modern liquid crystal displays (LCDs) for vision research. *Front Psychol,* 6**,** 303.

GIAVARINA, D. 2015. Understanding Bland Altman analysis. *Biochemia Medica,* 25**,** 141-151.

GIBSON, R. A. & SANDERSON, H. F. 1980. Observer variation in ophthalmology. *Br J Ophthalmol,* 64**,** 457-460.

GILBERT, M. 1953. Definition of Visual Acuity. *Br J Ophthalmol,* 37**,** 661-669.

GOH, Y. W., ANG, G. S. & AZUARA-BLANCO, A. 2011. Lifetime visual prognosis of patients with glaucoma. *Clin Experiment Ophthalmol,* 39**,** 766-770.

GONNERMANN, J., KLAMANN, M. K., MAIER, A. K., RJASANOW, J., JOUSSEN, A. M., BERTELMANN, E., RIECK, P. W. & TORUN, N. 2012. Visual outcome and complications after posterior iris-claw aphakic intraocular lens implantation. *J Cataract Refract Surg,* 38**,** 2139-2143.

GREEN, D. G. 1970. Regional variations in the visual acuity for interference fringes on the retina. *J Physiol,* 207**,** 351-356.

GREEN, J. 1868. On a New Series of Test-Letters for determining the Acuteness of Vision. *Trans Am Ophthalmol Soc,* 1**,** 68-71.

GRIMM, W., RASSOW, B., WESEMANN, W., SAUR, K. & HILZ, R. 1994. Correlation of optotypes with the Landolt ring--a fresh look at the comparability of optotypes. *Optom Vis Sci,* 71**,** 6-13.

HARRIS, S. J., HANSEN, R. M. & FULTON, A. B. 1984. Assessment of acuity in human infants using face and grating stimuli. *Invest Ophthalmol Vis Sci,* 25**,** 782-786.

HAZEL, C. A. & ELLIOTT, D. B. 2002. The dependency of logMAR visual acuity measurements on chart design and scoring rule. *Optom Vis Sci,* 79**,** 788-792.

HAZEL, C. A., LATHAM, K., ARMSTRONG, R. A., BENSON, M. T. & FROST, N. A. 2000. Visual Function and Subjective Quality of Life Compared in Subjects with Acquired Macular Disease. *Invest Ophthalmol Vis Sci,* 41**,** 1309-1315.

HEINRICH, S. P. & BACH, M. 2013. Resolution acuity versus recognition acuity with Landolt-style optotypes. *Graefes Arch Clin Exp Ophthalmol*, 251, 2235-2241.

HEINRICH, S. P., KRUGER, K. & BACH, M. 2010. The effect of optotype presentation duration on acuity estimates revisited. *Graefes Arch Clin Exp Ophthalmol,* 248**,** 389-394.

HESS, R. & WOO, G. 1978. Vision through cataracts. *Invest Ophthalmol Vis Sci,* 17**,** 428-435.

HESS, R. F. & CARNEY, L. G. 1979. Vision through an abnormal cornea: a pilot study of the relationship between visual loss from corneal distortion, corneal edema, keratoconus, and some allied corneal pathology. *Invest Ophthalmol Vis Sci,* 18**,** 476-483.

HOGG, R., CURRY, E., MULDREW, A., WINDER, J., STEVENSON, M., MCCLURE, M. & CHAKRAVARTHY, U. 2003. Identification of lesion components that influence visual function in age related macular degeneration. *Br J Ophthalmol,* 87**,** 609-614.

HOWARD, C. & FIRTH, A. Y. 2006. Is the Cardiff Acuity Test effective in detecting refractive errors in children? *Optom Vis Sci,* 83**,** 577-581.

HOWLAND, B., GINSBURG, A. & CAMPBELL, F. 1978. High-pass spatial frequency letters as clinical optotypes. *Vision Res,* 18**,** 1063-1066.

HUBEL, D. H. & WIESEL, T. N. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol,* 160**,** 106-154.

HUSSAIN, B., SALEH, G. M., SIVAPRASAD, S. & HAMMOND, C. J. 2006. Changing from Snellen to LogMAR: debate or delay? *Clin Experiment Ophthalmol,* 34**,** 6-8.

INTERNATIONAL COUNCIL OF OPHTHALMOLOGY 1988. Visual acuity measurement standard (1984). *Ital J Ophthalmol,* II**,** 1-15.

JACKSON, A. J. & BAILEY, I. L. 2004. Visual Acuity. *Optometry in Practice,* 5**,** 53-70.

JAVED, U., MCVEIGH, K., SCOTT, N. W. & AZUARA-BLANCO, A. 2015. Cataract extraction and patient vision-related quality of life: a cohort study. *Eye (Lond),* 29**,** 921-925.

JINABHAI, A., O'DONNELL, C., RADHAKRISHNAN, H. & NOURRIT, V. 2012. Forward light scatter and contrast sensitivity in keratoconic patients. *Cont Lens Anterior Eye,* 35**,** 22-27.

KAISER, P. K. 2009. Prospective evaluation of visual acuity assessment: a comparison of snellen versus ETDRS charts in clinical practice (An AOS Thesis). *Trans Am Ophthalmol Soc,* 107**,** 311-324.

KARA, S., GENCER, B., ERSAN, I., ARIKAN, S., KOCABIYIK, O., TUFAN, H. A. & COMEZ, A. 2016. Repeatability of contrast sensitivity testing in patients with age-related macular degeneration, glaucoma, and cataract. *Arquivos Brasileiros de Oftalmologia,* 79**,** 323-327.

KAY, H. 1983. New method of assessing visual acuity with pictures. *Br J Ophthalmol,* 67**,** 131-133.

KEANE, P. A., PATEL, P. J., LIAKOPOULOS, S., HEUSSEN, F. M., SADDA, S. R. & TUFAIL, A. 2012. Evaluation of age-related macular degeneration with optical coherence tomography. *Surv Ophthalmol,* 57**,** 389-414.

KLEIN, R., KLEIN, B. E. K., MOSS, S. E. & DEMETS, D. 1983. Inter-observer Variation in Refraction and Visual Acuity Measurement Using a Standardized Protocol. *Ophthalmology,* 90**,** 1357-1359.

KLEIN, R., WANG, Q., KLEIN, B. E., MOSS, S. E. & MEUER, S. M. 1995. The relationship of age-related maculopathy, cataract, and glaucoma to visual acuity. *Invest Ophthalmol Vis Sci,* 36**,** 182-191.

KLEINER, R. C., ENGER, C., ALEXANDER, M. F. & FINE, S. L. 1988. Contrast sensitivity in age-related macular degeneration. *Arch Ophthalmol,* 106**,** 55-57.

KNIESTEDT, C. & STAMPER, R. L. 2003. Visual acuity and its measurement. *Ophthalmol Clin North Am,* 16**,** 155-170.

KOENIG, S., TONAGEL, F., SCHIEFER, U., BACH, M. & HEINRICH, S. 2014. Assessing visual acuity across five disease types: ETDRS charts are faster with clinical outcome comparable to Landolt Cs. *Graefe's Arch Clin Exp Ophthalmol,* 252**,** 1093-1099.

KOLB, H., FERNANDEZ, E. & NELSON, R. 1995. *In:* KOLB, H., FERNANDEZ, E. & NELSON, R. (eds.) *Webvision: The Organization of the Retina and Visual System.* Salt Lake City (UT): University of Utah Health Sciences Center.

KOSKIN, S. A., BOIKO, E. V., SOBOLEV, A. F. & SHELEPIN, Y. E. 2007. Mechanisms of recognition of the outlines of "vanishing" optotypes. *Neurosci Behav Physiol,* 37**,** 59-65.

KUO, H.-K., KUO, M.-T., TIONG, I.-S., WU, P.-C., CHEN, Y.-J. & CHEN, C.-H. 2011. Visual acuity as measured with Landolt C chart and Early Treatment of Diabetic Retinopathy Study (ETDRS) chart. *Graefe's Arch Clin Exp Ophthalmol,,* 249**,** 601-605.

LAIDLAW, D. A., ABBOTT, A. & ROSSER, D. A. 2003. Development of a clinically feasible logMAR alternative to the Snellen chart: performance of the "compact reduced logMAR" visual acuity chart in amblyopic children. *Br J Ophthalmol,* 87**,** 1232-1234.

LAIDLAW, D. A., TAILOR, V., SHAH, N., ATAMIAN, S. & HARCOURT, C. 2008. Validation of a computerised logMAR visual acuity measurement system (COMPlog): comparison with ETDRS and the electronic ETDRS testing algorithm in adults and amblyopic children. *Br J Ophthalmol,* 92**,** 241-244.

LALWANI, G. A., ROSENFELD, P. J., FUNG, A. E., DUBOVY, S. R., MICHELS, S., FEUER, W., DAVIS, J. L., FLYNN, H. W., JR. & ESQUIABRO, M. 2009. A variable-dosing regimen with intravitreal ranibizumab for neovascular age-related macular degeneration: year 2 of the PrONTO Study. *Am J Ophthalmol,* 148**,** 43-58.e1.

LATHAM, K. 1998. Who uses contrast sensitivity in optometric practice? *Ophthalmic Physiol Opt,* 18**,** S2-S13.

LATHAM, K., KATSOU, M. F. & RAE, S. 2015. Advising patients on visual fitness to drive: implications of revised DVLA regulations. *Br J Ophthalmol,* 99**,** 545-548.

LATHAM, K. & USHERWOOD, C. 2010. Assessing visual activities of daily living in the visually impaired. *Ophthalmic Physiol Opt,* 30**,** 55-65.

LATHAM, K. & WHITAKER, D. 1996. A Comparison of Word Recognition and Reading Performance in Foveal and Peripheral Vision. *Vision Res,* 36**,** 2665-2674.

LEE, J. E., FOS, P. J., SUNG, J. H., AMY, B. W., ZUNIGA, M. A., LEE, W. J. & KIM, J. C. 2005. Relationship of cataract symptoms of preoperative patients and vision-related quality of life. *Qual Life Res,* 14**,** 1845-1853.

LEVI, D. M. 2008. Crowding--an essential bottleneck for object recognition: a mini-review. *Vision Res,* 48**,** 635-654.

LIM, L. A., FROST, N. A., POWELL, R. J. & HEWSON, P. 2010. Comparison of the ETDRS logMAR, 'compact reduced logMar' and Snellen charts in routine clinical practice. *Eye (Lond),* 24**,** 673-677.

LIU, L., WANG, Y. Z. & BEDELL, H. E. 2014. Visual-function tests for self-monitoring of age-related macular degeneration. *Optom Vis Sci,* 91**,** 956-965.

LOVIE-KITCHIN, J. E. 1988. Validity and reliability of visual acuity measurements. *Ophthalmic Physiol Opt,* 8**,** 363-370.

LOVIE-KITCHIN, J. E. 2015. Is it time to confine Snellen charts to the annals of history? *Ophthalmic Physiol Opt,* 35**,** 631-636.

LUU, C. D., DIMITROV, P. N., WU, Z., AYTON, L. N., MAKEYEVA, G., AUNG, K. Z., VARSAMIDIS, M., ROBMAN, L., VINGRYS, A. J. & GUYMER, R. H. 2013. Static and flicker perimetry in age-related macular degeneration. *Invest Ophthalmol Vis Sci,* 54**,** 3560-3568.

MACKIE, R. T., MCCULLOCH, D. L., SAUNDERS, K. J., BALLANTYNE, J., DAY, R. E., BRADNAM, M. S. & DUTTON, G. N. 1995. Comparison of visual assessment tests in multiply handicapped children. *Eye (Lond),* 9 (Pt 1)**,** 136-141.

MACULAR PHOTOCOAGULATION STUDY GROUP 1994. Visual outcome after laser photocoagulation for subfoveal choroidal neovascularization secondary to age-related macular degeneration. The influence of initial lesion size and initial visual acuity. Macular Photocoagulation Study Group. *Arch Ophthalmol,* 112**,** 480-488.

MAJAJ, N. J., PELLI, D. G., KURSHAN, P. & PALOMARES, M. 2002. The role of spatial frequency channels in letter identification. *Vision Res,* 42**,** 1165-1184.

MANNY, R. E., HUSSEIN, M., GWIAZDA, J. & MARSH-TOOTLE, W. 2003. Repeatability of ETDRS visual acuity in children. *Invest Ophthalmol Vis Sci,* 44**,** 3294-3300.

MATHEW, J. A., SHAH, S. A. & SIMON, J. W. 2011. Varying Difficulty of Snellen Letters and Common Errors in Amblyopic and Fellow Eyes. *Arch Ophthalmol,* 129**,** 184-187.

MCALINDEN, C., KHADKA, J. & PESUDOVS, K. 2011. Statistical methods for conducting agreement (comparison of clinical tests) and precision (repeatability or reproducibility) studies in optometry and ophthalmology. *Ophthalmic Physiol Opt,* 31**,** 330-338.

MCANANY, J. J., ALEXANDER, K. R., LIM, J. I. & SHAHIDI, M. 2011. Object frequency characteristics of visual acuity. *Invest Ophthalmol Vis Sci,* 52**,** 9534-9538.

MCGRAW, P., WINN, B. & WHITAKER, D. 1995. Reliability of the Snellen chart. *Bmj,* 310, 1481-1482.

MCMONNIES, C. W. 1999. Chart construction and letter legibility/readability. *Ophthalmic Physiol Opt,* 19, 498-506.

MCMONNIES, C. W. 2003. The dependency of LogMAR visual acuity measurements on chart design and scoring rule. *Optom Vis Sci,* 80, 486-487; author reply 487.

MCMONNIES, C. W. & HO, A. 2000. Letter legibility and chart equivalence. *Ophthalmic Physiol Opt,* 20, 142-152.

MEDINA, A. & HOWLAND, B. 1988. A novel high-frequency visual acuity chart. *Ophthalmic Physiol Opt,* 8, 14-18.

MICHAEL, R., VAN RIJN, L. J., VAN DEN BERG, T. J., BARRAQUER, R. I., GRABNER, G., WILHELM, H., COECKELBERGH, T., EMESZ, M., MARVAN, P. & NISCHLER, C. 2009. Association of lens opacities, intraocular straylight, contrast sensitivity and visual acuity in European drivers. *Acta Ophthalmol,* 87, 666-671.

MIDENA, E., DEGLI ANGELI, C., BLARZINO, M. C., VALENTI, M. & SEGATO, T. 1997. Macular function impairment in eyes with early age-related macular degeneration. *Invest Ophthalmol Vis Sci,* 38, 469-477.

MIDENA, E. & PILOTTO, E. 2017. Microperimetry in age: related macular degeneration. *Eye (Lond)*, 31, 985-994.

MINASSIAN, D., REIDY, A., DESAI, P., FARROW, S., VAFIDIS, G. & MINASSIAN, A. 2000. The deficit in cataract surgery in England and Wales and the escalating problem of visual impairment: epidemiological modelling of the population dynamics of cataract. *Br J Ophthalmol,* 84**,** 4-8.

MITCHELL, J. & BRADLEY, C. 2006. Quality of life in age-related macular degeneration: a review of the literature. *Health Qual Life Outcomes,* 4**,** 97.

MULHOLLAND, P. J., ZLATKOVA, M. B., REDMOND, T., GARWAY-HEATH, D. F. & ANDERSON, R. S. 2015. Effect of varying CRT refresh rate on the measurement of temporal summation. *Ophthalmic Physiol Opt,* 35**,** 582-590.

NAS-NRC COMMITTEE ON VISION 1980. Recommended stardard procedures for the clinical measurement and specification of visual acuity. Report of working group 39. *Adv Ophthalmol,* 41**,** 103-148.

NEELAM, K., NOLAN, J., CHAKRAVARTHY, U. & BEATTY, S. 2009. Psychophysical function in age-related maculopathy. *Surv Ophthalmol,* 54**,** 167-210.

NESTARES, O., NAVARRO, R. & ANTONA, B. 2003. Bayesian model of Snellen visual acuity. *J Opt Soc Am A Opt Image Sci Vis,* 20**,** 1371-1381.

NIWA, K. & TOKORO, T. 1997. Measurement of temporal summation of visual acuity with use of modified tachistoscope. *Jpn J Ophthalmol,* 41**,** 403-408.

OWSLEY, C. 2011. Aging and Vision. *Vision Res,* 51**,** 1610-1622.

PARDHAN, S. 1997. Crowding in visually impaired patients: contour interaction and/or gaze-selection defects? *J Neuroophthalmol,* 18**,** 59-65.

PARISH, D. H. & SPERLING, G. 1991. Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Res,* 31**,** 1399-1415.

PATEL, P. J., CHEN, F. K., RUBIN, G. S. & TUFAIL, A. 2008. Intersession repeatability of visual acuity scores in age-related macular degeneration. *Invest Ophthalmol Vis Sci,* 49**,** 4347-4352.

PAUDEL, N., JACOBS, R. J., SLOAN, R., DENNY, S., SHEA, K., THOMPSON, B. & ANSTICE, N. 2017. Effect of simulated refractive error on adult visual acuity for paediatric tests. *Ophthalmic Physiol Opt,* 37**,** 521-530.

PELLI, D. G. 1997. Pixel independence: measuring spatial interactions on a CRT display. *Spat Vis,* 10**,** 443-446.

PHIPPS, J. A., DANG, T. M., VINGRYS, A. J. & GUYMER, R. H. 2004. Flicker perimetry losses in age-related macular degeneration. *Invest Ophthalmol Vis Sci,* 45**,** 3355-3360.

PLAINIS, S., KONTADAKIS, G., FELONI, E., GIANNAKOPOULOU, T., TSILIMBARIS, M. K., PALLIKARIS, I. G. & MOSCHANDREAS, J. 2013. Comparison of visual acuity charts in young adults and patients with diabetic retinopathy. *Optom Vis Sci,* 90**,** 174-178.

POULERE, E., MOSCHANDREAS, J., KONTADAKIS, G. A., PALLIKARIS, I. G. & PLAINIS, S. 2013. Effect of blur and subsequent adaptation on visual acuity using letter and Landolt C charts: differences between emmetropes and myopes. *Ophthalmic Physiol Opt,* 33**,** 130-137.

PRINCE, J. H. & FRY, G. A. 1958. Correction for the guessing bias in the Landolt ring test. *Am J Optom Arch Am Acad Optom,* 35**,** 134-141.

QIU, F. & LEAT, S. J. 2009. Functional deficits in early stage age-related maculopathy. *Clin Exp Optom,* 92**,** 90-98.

RAASCH, T. W., BAILEY, I. L. & BULLIMORE, M. A. 1998. Repeatability of visual acuity measurement. *Optom Vis Sci,* 75**,** 342-348.

RAVIKUMAR, S., BRADLEY, A. & THIBOS, L. 2010. Phase changes induced by optical aberrations degrade letter and face acuity. *J Vis,* 10**,** 18.

RAZALI, N. M. & WAH, Y. B. 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics,* 2**,** 21-33.

REDMOND, T. 2012. Toward the optimum measurement of visual acuity. *Invest Ophthalmol Vis Sci,* 53**,** 7424.

REEVES, B. C., WOOD, J. M. & HILL, A. R. 1991. Vistech VCTS 6500 charts--within- and between-session reliability. *Optom Vis Sci,* 68**,** 728-737.

REEVES, B. C., WOOD, J. M. & HILL, A. R. 1993. Reliability of high- and low-contrast letter charts. *Ophthalmic Physiol Opt,* 13**,** 17-26.

REICH, L. N. & BEDELL, H. E. 2000. Relative legibility and confusions of letter acuity targets in the peripheral and central retina. *Optom Vis Sci,* 77**,** 270-275.

RICCI, F., CEDRONE, C. & CERULLI, L. 1998. Standardized measurement of visual acuity. *Ophthalmic Epidemiol,* 5**,** 41-53.

ROCHA, K. M., NOSE, W., BOTTOS, K., BOTTOS, J., MORIMOTO, L. & SORIANO, E. 2007. Higher-order aberrations of age-related cataract. *J Cataract Refract Surg,* 33**,** 1442-1446.

ROFAGHA, S., BHISITKUL, R. B., BOYER, D. S., SADDA, S. R. & ZHANG, K. 2013. Seven-year outcomes in ranibizumab-treated patients in ANCHOR, MARINA, and

HORIZON: a multicenter cohort study (SEVEN-UP). *Ophthalmology,* 120**,** 2292-2299.

ROSSER, D. A., COUSENS, S. N., MURDOCH, I. E., FITZKE, F. W. & LAIDLAW, D. A. 2003a. How sensitive to clinical change are ETDRS logMAR visual acuity measurements? *Invest Ophthalmol Vis Sci,* 44**,** 3278-3281.

ROSSER, D. A., LAIDLAW, D. A. & MURDOCH, I. E. 2001. The development of a "reduced logMAR" visual acuity chart for use in routine clinical practice. *Br J Ophthalmol,* 85**,** 432-436.

ROSSER, D. A., MURDOCH, I. E. & COUSENS, S. N. 2004. The effect of optical defocus on the test-retest variability of visual acuity measurements. *Invest Ophthalmol Vis Sci,* 45**,** 1076-1079.

ROSSER, D. A., MURDOCH, I. E., FITZKE, F. W. & LAIDLAW, D. A. 2003b. Improving on ETDRS acuities: design and results for a computerised thresholding device. *Eye (Lond),* 17**,** 701-706.

RUBIN, G. S. 2013. Measuring reading performance. *Vision Res,* 90**,** 43-51.

RUBIN, G. S., ADAMSONS, I. A. & STARK, W. J. 1993. Comparison of acuity, contrast sensitivity, and disability glare before and after cataract surgery. *Arch Ophthalmol,* 111**,** 56-61.

RUBIN, G. S. & FEELY, M. 2009. The Role of Eye Movements During Reading in Patients with Age-Related Macular Degeneration (AMD). *J Neuroophthalmol,* 33**,** 120-126.

RUNGE, P. E. 2000. Eduard Jaeger's Test-Types (Schrift-Scalen) and the historical development of vision tests. *Trans Am Ophthalmol Soc,* 98**,** 375-438.

SCHWIEGERLING, J. 2000. Theoretical limits to visual performance. *Surv Ophthalmol,* 45**,** 139-146.

SEIPLE, W., HOLOPIGIAN, K., SHNAYDER, Y. & SZLYK, J. P. 2001. Duration thresholds for target detection and identification in the peripheral visual field. *Optom Vis Sci,* 78**,** 169-176.

SEIPLE, W., HOLOPIGIAN, K., SZLYK, J. P. & GREENSTEIN, V. C. 1995. The effects of random element loss on letter identification: Implications for visual acuity loss in patients with retinitis pigmentosa. *Vision Res,* 35**,** 2057-2066.

SHAH, N., DAKIN, S. C. & ANDERSON, R. S. 2012a. Effect of optical defocus on detection and recognition of vanishing optotype letters in the fovea and periphery. *Invest Ophthalmol Vis Sci,* 53**,** 7063-7070.

SHAH, N., DAKIN, S. C. & ANDERSON, R. S. 2016a. Visual acuity - more than meets the eye? *Optometry Today,(May)***,** 69-73.

SHAH, N., DAKIN, S. C., DOBINSON, S., TUFAIL, A., EGAN, C. A. & ANDERSON, R. S. 2016b. Visual acuity loss in patients with age-related macular degeneration measured using a novel high-pass letter chart. *Br J Ophthalmol,* 100**,** 1346-1352.

SHAH, N., DAKIN, S. C., REDMOND, T. & ANDERSON, R. S. 2011a. Vanishing Optotype acuity: repeatability and effect of the number of alternatives. *Ophthalmic Physiol Opt,* 31**,** 17-22.

SHAH, N., DAKIN, S. C., WHITAKER, H. L. & ANDERSON, R. S. 2014. Effect of scoring and termination rules on test-retest variability of a novel high-pass letter acuity chart. *Invest Ophthalmol Vis Sci,* 55**,** 1386-1392.

SHAH, N., LAIDLAW, D. A., BROWN, G. & ROBSON, C. 2010. Effect of letter separation on computerised visual acuity measurements: comparison with the gold standard Early Treatment Diabetic Retinopathy Study (ETDRS) chart. *Ophthalmic Physiol Opt,* 30**,** 200-203.

SHAH, N., LAIDLAW, D. A., RASHID, S. & HYSI, P. 2012b. Validation of printed and computerised crowded Kay picture logMAR tests against gold standard ETDRS acuity test chart measurements in adult and amblyopic paediatric subjects. *Eye (Lond),* 26**,** 593-600.

SHAH, N., LAIDLAW, D. A., SHAH, S. P., SIVASUBRAMANIAM, S., BUNCE, C. & COUSENS, S. 2011b. Computerized repeating and averaging improve the test-retest variability of ETDRS visual acuity measurements: implications for sensitivity and specificity. *Invest Ophthalmol Vis Sci,* 52**,** 9397-9402.

SHAMIR, R. R., FRIEDMAN, Y. G., JOSKOWICZ, L., MIMOUNI, M. & BLUMENTHAL, E. Z. 2016. The influence of varying the number of characters per row on the accuracy and reproducibility of the ETDRS visual acuity chart. *Graefes Arch Clin Exp Ophthalmol*, 254, 971-976.

SHEEDY, J. E., BAILEY, I. L. & RAASCH, T. W. 1984. Visual acuity and chart luminance. *Am J Optom Physiol Opt,* 61**,** 595-600.

SHONA, O., GUPTA, B., VEMALA, R. & SIVAPRASAD, S. 2011. Visual acuity outcomes in ranibizumab-treated neovascular age-related macular degeneration; stratified by baseline vision. *Clin Experiment Ophthalmol,* 39**,** 5-8.

SLOAN, L. L., ROWLAND, W. M. & ALTMAN, A. 1952. Comparison of three types of test target for the measurement of visual acuity. *Q Rev Opthalmol,* 8**,** 4-16.

SMITH, M. 2005. Does encouragement boost visual acuity testing results? *Nurs Times,* 101**,** 38-41.

SNELLEN, H. 1864. Test types for the determination of acuteness of vision. *Official war office edition (2nd edn).* Utrecht.

SOLOMON, J. A. & PELLI, D. G. 1994. The visual filter mediating letter identification. *Nature,* 369**,** 395-397.

STANGOS, N., VOUTAS, S., TOPOUZIS, F. & KARAMPATAKIS, V. 1995. Contrast sensitivity evaluation in eyes predisposed to age-related macular degeneration and presenting normal visual acuity. *Ophthalmologica,* 209**,** 194-198.

STEINBERG, E. P., TIELSCH, J. M., SCHEIN, O. D., JAVITT, J. C., SHARKEY, P., CASSARD, S. D., LEGRO, M. W., DIENER-WEST, M., BASS, E. B., DAMIANO, A. M. & ET AL. 1994. The VF-14. An index of functional impairment in patients with cataract. *Arch Ophthalmol,* 112**,** 630-638.

STEWART, C. E., HUSSEY, A., DAVIES, N. & MOSELEY, M. J. 2006. Comparison of logMAR ETDRS chart and a new computerised staircased procedure for assessment of the visual acuity of children. *Ophthalmic Physiol Opt,* 26**,** 597-601.

SUPERSTEIN, R., BOYANER, D., OVERBURY, O. & COLLIN, C. 1997. Glare disability and contrast sensitivity before and after cataract surgery. *J Cataract Refract Surg,* 23**,** 248-253.

TEWARI, H. K., KORI, V., SONY, P., VENKATESH, P. & GARG, S. 2006. Snellen chart may be preferable over early treatment diabetic retinopathy study charts for rapid visual acuity assessment. *Indian J Ophthalmol,* 54**,** 214.

THAYAPARAN, K., CROSSLAND, M. D. & RUBIN, G. S. 2007. Clinical assessment of two new contrast sensitivity charts. *Br J Ophthalmol,* 91**,** 749-752.

THE ROYAL COLLEGE OF OPHTHALMOLOGISTS. 2015. *Commissioning Guide: Cataract Surgery* [Online]. Available: https://www.rcophth.ac.uk/wp-content/uploads/2015/03/Commissioning-Guide-Cataract-Surgery-Final-February-2015.pdf [Accessed 12/01/2017].

THIBOS, L. N. 1998. Acuity perimetry and the sampling theory of visual resolution. *Optom Vis Sci,* 75**,** 399-406.

THIBOS, L. N. & BRADLEY, A. 1993. New methods for discriminating neural and optical losses of vision. *Optom Vis Sci,* 70**,** 279-287.

THIBOS, L. N., CHENEY, F. E. & WALSH, D. J. 1987a. Retinal limits to the detection and resolution of gratings. *J Opt Soc Am A,* 4**,** 1524-1529.

THIBOS, L. N., STILL, D. L. & BRADLEY, A. 1996. Characterization of spatial aliasing and contrast sensitivity in peripheral vision. *Vision Res,* 36**,** 249-258.

THIBOS, L. N., WALSH, D. J. & CHENEY, F. E. 1987b. Vision beyond the resolution limit: aliasing in the periphery. *Vision Res,* 27**,** 2193-2197.

THOMSON, D. 2005a. VA testing in optometric practice part 1: the Snellen chart. *Optometry Today,(April)***,** 56-58.

THOMSON, D. 2005b. VA testing in optometric practice. Part 2: Newer chart designs. *Optometry Today,* 45**,** 22-24.

THORN, F. & SCHWARTZ, F. 1990. Effects of dioptric blur on Snellen and grating acuity. *Optom Vis Sci,* 67**,** 3-7.

TREACY, M. P., HURST, T. P., CONWAY, M., DUIGNAN, E. S., DIMITROV, B. D., BRENNAN, N. & CASSIDY, L. 2015. The Early Treatment in Diabetic Retinopathy Study Chart Compared with the Tumbling-E and Landolt-C. *Ophthalmology,* 122**,** 1062-1063.

VAN DEN BERG, T. J., VAN RIJN, L. J., MICHAEL, R., HEINE, C., COECKELBERGH, T., NISCHLER, C., WILHELM, H., GRABNER, G., EMESZ, M., BARRAQUER, R. I., COPPENS, J. E. & FRANSSEN, L. 2007. Straylight effects with aging and lens extraction. *Am J Ophthalmol,* 144**,** 358-363.

VAN DER MEULEN, I. J., GJERTSEN, J., KRUIJT, B., WITMER, J. P., RULO, A., SCHLINGEMANN, R. O. & VAN DEN BERG, T. J. 2012. Straylight measurements as an indication for cataract surgery. *J Cataract Refract Surg,* 38**,** 840-848.

VANDEN BOSCH, M. E. & WALL, M. 1997. Visual acuity scored by the letter-by-letter or probit methods has lower retest variability than the line assignment method. *Eye (Lond),* 11 (Pt 3)**,** 411-417.

WALKER, J. P. S. 1942. Test Type. *Br J Ophthalmol,* 26**,** 556-559.

WALL, M., CHAUHAN, B., FRISEN, L., HOUSE, P. H. & BRITO, C. 2004. Visual field of high-pass resolution perimetry in normal subjects. *J Glaucoma,* 13**,** 15-21.

WANG, P. & NIKOLIC, D. 2011. An LCD Monitor with Sufficiently Precise Timing for Research in Vision. *Front Hum Neurosci,* 5**,** 1-10.

WANG, Y. Z., THIBOS, L. N. & BRADLEY, A. 1997. Effects of refractive error on detection acuity and resolution acuity in peripheral vision. *Invest Ophthalmol Vis Sci,* 38**,** 2134-2143.

WATSON, A. B. & AHUMADA, A. J. 2012. Modeling acuity for optotypes varying in complexity. *J Vis,* 12, 1-19.

WATSON, A. B. & AHUMADA, A. J. 2015. Letter identification and the Neural Image Classifier. *J Vis,* 15**,** 1-26.

WATSON, A. B. & PELLI, D. G. 1983. QUEST: a Bayesian adaptive psychometric method. *Percept Psychophys,* 33**,** 113-120.

WESTHEIMER, G. 1965. VISUAL ACUITY. *Annu Rev Psychol,* 16**,** 359-380.

WHITNEY, D. & LEVI, D. M. 2011. Visual Crowding: a fundamental limit on conscious perception and object recognition. *Trends Cogn Sci,* 15**,** 160-168.

WILLIAMS, D. R. & COLETTA, N. J. 1987. Cone spacing and the visual resolution limit. *J Opt Soc Am A,* 4**,** 1514-1523.

WITTICH, W., OVERBURY, O., KAPUSTA, M. A. & WATANABE, D. H. 2006. Differences between recognition and resolution acuity in patients undergoing macular hole surgery. *Invest Ophthalmol Vis Sci,* 47**,** 3690-3694.

WOLFFSOHN, J. S., ANDERSON, S. J., MITCHELL, J., WOODCOCK, A., RUBINSTEIN, M., FFYTCHE, T., BROWNING, A., WILLBOND, K., AMOAKU, W. M. & BRADLEY, C. 2006. Effect of age related macular degeneration on the Eger macular stressometer photostress recovery time. *Br J Ophthalmol,* 90**,** 432-434.

WONG, T. Y., CHAKRAVARTHY, U., KLEIN, R., MITCHELL, P., ZLATEVA, G., BUGGAGE, R., FAHRBACH, K., PROBST, C. & SLEDGE, I. 2008. The natural history and prognosis of neovascular age-related macular degeneration: a systematic review of the literature and meta-analysis. *Ophthalmology,* 115**,** 116-126.

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 Glob Health,* 2**,** e106-e116.

WOODHOUSE, J. M., ADOH, T. O., ODUWAIYE, K. A., BATCHELOR, B. G., MEGJI, S., UNWIN, N. & JONES, N. 1992. New acuity test for toddlers. *Ophthalmic Physiol Opt,* 12**,** 249-251.

WU, Z., AYTON, L. N., GUYMER, R. H. & LUU, C. D. 2013. Intrasession test-retest variability of microperimetry in age-related macular degeneration. *Invest Ophthalmol Vis Sci,* 54**,** 7378-7385.

ZAJAC-PYTRUS, H. M., PILECKA, A., TURNO-KRECICKA, A., ADAMIEC-MROCZEK, J. & MISIUK-HOJLO, M. 2015. The Dry Form of Age-Related Macular Degeneration (AMD): The Current Concepts of Pathogenesis and Prospects for Treatment. *Adv Clin Exp Med,* 24**,** 1099-1104.

ZHU, X., YE, H., HE, W., YANG, J., DAI, J. & LU, Y. 2016. Objective functional visual outcomes of cataract surgery in patients with good preoperative visual acuity. *Eye*, 31, 452-459.

ZLATKOVA, M. B., COULTER, E. E. & ANDERSON, R. S. 2006. The effect of simulated lens yellowing and opacification on blue-on-yellow acuity and contrast sensitivity. *Vision Res,* 46**,** 2432-2442.

Appendix A – Publications

A

Peer-reviewed publications originating from this thesis

SHAH, N., DAKIN, S. C., REDMOND, T. & ANDERSON, R. S. 2011. Vanishing Optotype acuity: repeatability and effect of the number of alternatives. *Ophthalmic Physiol Opt,* 31**,** 17-22.

SHAH, N., DAKIN, S. C. & ANDERSON, R. S. 2012. Effect of optical defocus on detection and recognition of vanishing optotype letters in the fovea and periphery. *Invest Ophthalmol Vis Sci,* 53**,** 7063-70.

(Received Research Highlight: REDMOND, T., 2012. Toward the optimum measurement of visual acuity. *Invest Ophthalmol Vis Sci,* 53, 7424).

SHAH, N., DAKIN, S. C., WHITAKER, H. L. & ANDERSON, R. S. 2014. Effect of scoring and termination rules on test-retest variability of a novel high-pass letter acuity chart. *Invest Ophthalmol Vis Sci,* 55**,** 1386-92.

SHAH, N., DAKIN, S. C., DOBINSON, S., TUFAIL, A., EGAN, C. A. & ANDERSON, R. S. 2016. Visual acuity loss in patients with age-related macular degeneration measured using a novel high-pass letter chart. *Br J Ophthalmol*, 100(10), 1346-1352.

Additional publications

SHAH, N., DAKIN, S. C. & ANDERSON, R. S. 2016. Visual acuity - more than meets the eye? *Optometry Today,* 69-73.

ANDERSON, R.S. & **SHAH, N.** 2016. More sensitive visual acuity test for age-related macular degeneration. *Eye News,* 23(3), 46-50.

Appendix B - Presentations

B

Conference Presentations

2010 Hospital Optometrists Annual Conference (HOAC 36), Heathrow, UK (Paper presentation).

2010 UK and Eire Glaucoma Society (UKEGS), Westminster, UK (Poster presentation).

2011 Association for Research in Vision and Ophthalmology (ARVO), Fort Lauderdale, USA (Poster presentation).

2012 European Association of Optometry and Optics Conference (EAOO), Dublin, Ireland (Paper presentation).

2012 British Congress of Optometry and Vision Science (BCOVS), Bradford, UK (Paper presentation).

2012 Hospital Optometrists Annual Conference (HOAC 38), Chester, UK (Poster presentation).

2013 European Association of Optometry and Optics Conference (EAOO), Malaga, Spain (Poster presentation).

2013 Association for Research in Vision and Ophthalmology (ARVO), Seattle, USA (Poster presentation).

2013 British Congress of Optometry and Vision Science (BCOVS), Glasgow, Scotland, UK (Poster presentation).

2014 European Association of Optometry and Optics Conference (EAOO), Warsaw, Poland (Paper presentation).

2015 European Society for Low Vision Research and Rehabilitation (ESLRR), Oxford, UK (Poster presentation).

Appendix C – Press Releases

C

29/03/2016 Fight For Sight

http://www.fightforsight.org.uk/news-and-views/articles/news/new-sight-test-detects-early-amd/

30/03/2016 Association of Optometrists

https://www.aop.org.uk/ot/science-and-vision/research/2016/03/30/vanishing-letters-to-diagnose-amd-sooner

05/04/2016 RNIB

http://www.rnib.org.uk/nb-online/new-AMD-test

27/06/2016 Moorfields Eye Charity

https://moorfieldseyecharity.org.uk/news/new-test-detects-early-signs-amd?gclid=EAIaIQobChMIit_bxsnX1QIVT7HtCh0wMARpEAAYASAAEgITkPD_BwE