Optimisation of perimetric stimuli for mapping changes in spatial summation in glaucoma

A thesis submitted to Cardiff University for the degree of Doctor of Philosophy

By

Lindsay Catherine Rountree

School of Optometry and Vision Sciences
Cardiff University

April 2018

Supervised by Dr Tony Redmond, Dr Pádraig J. Mulholland and Prof. Roger S. Anderson

Acknowledgements

I applied for this PhD because I wanted a new challenge, and it has certainly been that! There have been tremendous highs, and deep lows, but I have never regretted it. Thank you to all at Cardiff University for the opportunity, with a particular mention to some key characters, without whom this would not have been possible.

I would like to thank my supervisors, Dr Tony Redmond, Dr Pádraig Mulholland, and Prof. Roger Anderson, and my advisor Prof. Rachel North, for all your guidance, support and encouragement. Particular thanks to Tony for pushing me, often outside my comfort zone, to be a better researcher – despite some occasional complaining, I really am grateful. Thank you also to The College of Optometrists for funding this research, and providing me with this opportunity.

A huge thank you to everyone who participated in this study. You all committed a lot of time and effort, and it is immensely appreciated. Thank you to Georgina Butler and Vikki Baker for your assistance with recruitment, to Katherine Ward who helped collect data on the effects of optical defocus (chapter six) as part of her undergraduate research project, and to Prof. James Morgan and Prof. David Garway-Heath for their input into the findings of the experiment presented in chapter four.

Thank you to all my fellow PhD colleagues for your friendship, encouragement and support, to those who have shared an office with me, and particularly to Louise, Shindy, Nikki, Flors, and Aysha, who have been there for the highs and lows. I would also like to thank the office staff for all their help, and in particular, of course, Sue Hobbs – without you none of us would ever finish!

Thank you to my family, and particularly my parents, Richard and Iris Rountree, who have supported and encouraged me throughout all my endeavours, academic and otherwise.

Finally, the biggest thank you of all goes to my husband, Toby Jackson, without whom I would never have had the courage to even apply for this PhD. For your never-ending support, and unwavering belief in my abilities, I am forever grateful.

Summary

Despite being considered the current reference standard for perimetric testing in glaucoma, standard automated perimetry has several cardinal limitations, including an unacceptably high test-retest variability, which increases with increasing depth of defect, and a limited useable dynamic range, with test-retest variability spanning almost the entire instrument range in advanced glaucomatous damage.

Prior studies have shown that spatial summation, the mechanism by which the visual system integrates light energy across the area of a stimulus, differs in disease, with an enlarged Ricco's area (the limit of complete spatial summation) found in individuals with glaucoma. The aim of this work was to investigate whether a perimetric stimulus designed to exploit these changes in spatial summation would enable a greater signal/noise ratio (SNR) than that of the current standard stimulus, by directly measuring the displacement of the spatial summation function in glaucoma. Three stimulus forms were developed; one varying in area alone, one varying in both area and contrast simultaneously, and one varying in contrast alone, all operating within the local Ricco's area. These novel stimuli were compared with the standard Goldmann III stimulus, in terms of disease signal, noise, and SNR.

The experiments presented in this thesis indicate that a stimulus modulating in area alone may offer greater benefits for measuring glaucomatous changes in spatial summation in a clinical setting, in the form of a greater disease signal, more uniform response variability with depth of defect, and greater SNR, when compared with the standard Goldmann III stimulus. Additionally, there is some indication that this stimulus is more robust to the effects of intraocular straylight than the Goldmann III stimulus, although test-retest variability and robustness to optical defocus are largely similar.

As this work represents the early investigations of this stimulus, further work is required to examine its translation into a clinical environment.

Table of Contents

Chapter 2 Selection and characterisation of an appropriate display screen

Chapter 3 The psychometric function, and investigation of response variability characteristics with Goldmann I-V

Chapter 4 Quantifying the signal/noise ratio with perimetric stimuli optimised to probe changing spatial summation in glaucoma

Chapter 5 Test-retest variability of perimetric stimuli optimised to probe changing spatial summation in glaucoma

Chapter 6 Resistance of perimetric stimuli, optimised to probe changing spatial summation in glaucoma, to optical defocus and intraocular straylight

Chapter 7 Overall discussion and future work

Table of Figures

Chapter 3 The psychometric function, and investigation of response variability characteristics with Goldmann I-V

Chapter 4 Quantifying the signal/noise ratio with perimetric stimuli optimised to probe changing spatial summation in glaucoma

Chapter 5 Test-retest variability of perimetric stimuli optimised to probe changing spatial summation in glaucoma

Chapter 6 Resistance of perimetric stimuli, optimised to probe changing spatial summation in glaucoma, to optical defocus and intraocular straylight

Abbreviations and Acronyms

2AFC	Two Alternative Forced Choice
A	Stimulus modulating in area only
AC	Stimulus modulating in both area and contrast simultaneously
ANOVA	Analysis of Variance
cd	Candelas
CFF	Critical Flicker Frequency
cm	Centimetres
C_R	Stimulus modulating in contrast only, within Ricco's area
CRS	Cambridge Research Systems
CRT	Cathode Ray Tube
CSS	Complete Spatial Summation
D	Dioptres
dB	Decibels
Deg	Degrees of visual angle
DC	Dioptre Cylinder
DS	Dioptre Sphere
e.g.	For Example
EGS	European Glaucoma Society
ERG	Electroretinogram
ESD	Estimated Standard Deviation (C-Quant)
FDT	Frequency Doubling Technology
fMRI	Functional Magnetic Resonance Imaging
FOS	Frequency-Of-Seeing
FPS	Frames Per Second
GHT	Glaucoma Hemifield Test
GIII	Goldmann III-equivalent stimulus
GPA	Guided Progression Analysis
GRP	Grating Resolution Perimetry
HEP	Heidelberg Edge Perimeter
HFA	Humphrey Field Analyzer
HUMO	Highest Unoccupied Molecular Orbital
Hz	Hertz

i.e.	Id est ('that is')
IOP	Intraocular Pressure
IQR	Interquartile Range
IS_O	Initial Stimulus (Original)
IS_S	Initial Stimulus (Supplementary)
ISS	Incomplete Spatial Summation
LCD	Liquid Crystal Display
LED	Light Emitting Diode
LGN	Lateral Geniculate Nucleus
LM	Mean Luminance
LU	Luminance Uniformity
LUMO	Lowest Unoccupied Molecular Orbital
LUT	Look-Up Table
MD	Mean Deviation
mm	Millimetres
Moorfields MDT	Moorfields Motion Displacement Test
MOCS	Method of Constant Stimuli
NICE	National Institute for Health and Care Excellence
NTG	Normal Tension Glaucoma
OCT	Optical Coherence Tomography
OHT	Ocular Hypertension
OLED	Organic Light Emitting Diode
OLS	Ordinary Least Squares
PERG	Pattern Electroretinogram
POAG	Primary Open Angle Glaucoma
PoPLR	Permutation of Pointwise Linear Regression
Q	Quality factor (C-Quant)
RGB	Red-Green-Blue
RGC	Retinal Ganglion Cell
RNU	Range of Non-Uniformity
SAP	Standard Automated Perimetry
ScM	Mean Luminance Output across display
SD	Standard Deviation
SE	Standard Error
SITA	Swedish Interactive Thresholding Algorithm

SNR	Signal/Noise Ratio
STP	Size Threshold Perimetry
SWAP	Short Wavelength Automated Perimetry
TD	Total Deviation
TD$_{SAP}$	Total Deviation from the SITA Standard 24-2 strategy (HFA II)
ZEST	Zippy Estimation of Sequential Testing

Chapter 1 Introduction and literature review

1.1   Introduction

Evaluation of the visual field is an important component of ocular examination, and Standard Automated Perimetry (SAP) is a well-established and commonly used clinical test, both in high-street optometric practice and in the hospital eye service. It can be used to aid in the identification of a wide range of ocular and neurological disorders, and is useful for formal certification of visual function. Sequential examinations are often used to follow the course of various conditions, and to monitor the effectiveness of treatment (Scheifer et al. 2005). Much of its use is in the detection and management of glaucoma, the primary focus of the study presented in this thesis, although its limitations in this role are well documented. Much of its development as a clinical technique occurred while little of the aetiology and pathological process of glaucoma was understood, which may, in part, explain why the technique is subject to these limitations. Perimetry may be expected to continually evolve as the multifactorial process of glaucoma development becomes increasingly understood.

Perimetry perhaps plays an underrated role in the clinical investigation of glaucoma. While other investigative techniques estimate the structural damage caused by glaucoma, perimetry estimates the functional effects of this damage, aiding not only in the clinical investigation of the disease, but also in understanding its impact on visual perception and, increasingly, quality of life.

The present chapter provides a general introduction to the glaucomatous process, and the continued development of perimetry as this disease process has become more readily understood. It details the limitations that still occur with this technique, and some of the more recent attempts to address them, with varying degrees of success. Finally, the chapter discusses the phenomenon of spatial summation, and the potential this may play in further perimetric development.

1.2 Glaucoma

1.2.1 A brief history

It is often reported that the first mention of glaucoma occurred in Hippocratic writings, as a blinding disease ('Glaucosis') in the elderly (Tsatsos and Broadway 2007). The description that is given suggests acute, symptomatic glaucoma, and it is likely that the associated corneal oedema resulted in some confusion between it and other conditions. It is unclear when the distinction between different types of glaucoma began, but an evaluation of the condition in 1858 listed two distinct 'long recognised' categories of 'acute' and 'chronic' glaucoma, categories still in use today (Hulke 1858). Indeed from this early paper, a familiar description of both acute and chronic glaucoma can be found, with the recognition of 'internal pressure' and 'hardness of the globe' as key features.

Although the Helmholtz Ophthalmoscope had only been made available seven years prior, the 'excavated state' of the optic disc is described for both acute and chronic glaucoma in this paper, likely informed by histological studies. Hulke (1858) credits von Graefe as the person who first identified this optic disc excavation, and who first 'called attention to the diminished size of the field of vision'. Von Graefe is also credited with introducing visual field testing and perimetry to the clinical setting (Johnson et al. 2011), illustrating that the development of the perimetric technique occurred while the glaucomatous disease process was largely unknown.

Hulke's description of glaucoma (1858) focused on the acute form, which appeared to be better defined at the time than the chronic form, and noted that the observation of both chronic and acute glaucoma in the same person led to the conclusion that they were two forms of the same disease. Although better defined, the description of acute glaucoma includes some retinal features that would not be recognised now as part of the glaucomatous process, indicating that the condition had not been fully differentiated from other retinal diseases. Indeed, the official definition of glaucoma is still updated periodically as knowledge of the condition is gained.

1.2.2   Definition of glaucoma

Reports from the 1930s (Pickard 1931), suggested the presence of glaucoma in the absence of an elevated intraocular pressure (IOP), now termed 'normal tension glaucoma' (NTG). Shields and Spaeth (2012), in describing the glaucoma definitions of the 20th century, illustrate the subsequent shift away from the emphasis on elevated IOP as a key feature of the condition (1950s), to the description of a 'characteristic optic neuropathy', with raised IOP as a risk factor rather than the primary cause. Shields and Spaeth (2012) report that subsequent questions were raised over whether IOP played any real role in glaucoma at all, although this has now been confirmed. Indeed, glaucoma is often induced in animal models by raising IOP experimentally (Urcola et al. 2006; Morrison et al. 2008; Morgan and Tribble 2015), demonstrating that an elevated IOP alone is sufficient to cause glaucoma, and IOP control is still considered the only proven method of treatment available (Weinreb et al. 2014). Investigations of other techniques, for example 'neuroprotective' therapies that maintain the health of the retinal ganglion cells, or 'neuroregenerative' therapies that promote cell repair or replacement (Kolko 2015; Shen et al. 2015; Song et al. 2015; Almasieh and Levin 2017), are currently under investigation.

The current definition has expanded still further, describing glaucoma as 'a group of progressive optic neuropathies characterised by degeneration of retinal ganglion cells and resulting changes in the optic nerve head. Loss of ganglion cells is related to the level of IOP, but other factors may also play a role' (Weinreb et al. 2014). This definition identifies many, more recent discoveries about glaucoma. For example, it is a 'group' of diseases, and in addition to the 'acute' and 'chronic' categories recognised in the 19th century, glaucoma may also be referred to as 'primary' or 'secondary', and 'open' or 'closed' angle, in addition to the many sub-types that are currently recognised (Casson et al. 2012). The retinal ganglion cells, whose axons form the optic nerve, are recognised as the target cells of the disease, and finally, 'ocular hypertension' (OHT), in which there is a raised IOP but no evidence of damage, is now differentiated as separate from true glaucoma.

The study presented in this thesis is fundamentally concerned with 'primary open angle glaucoma' (POAG), the most common form of glaucoma, and one which presents many challenges in its identification and continued management.

1.2.3 Pathogenesis of glaucoma

The pathogenesis of glaucoma is a continuing field of research, as it is not yet fully understood. It is widely accepted that glaucoma cannot be explained by a single mechanism, not least because there are so many forms of the disease. Even within one sub-category of glaucoma, for example POAG, there are great variations in susceptibility and disease progression, and as such the aetiology is accepted to be multifactorial (Fechtner and Weinreb 1994). Differing theories tended to fit into two broad categories, mechanical and vasogenic, although more recent studies suggest it likely that both factors are at play. Although a detailed analysis of the pathogenesis of glaucoma is beyond the scope of this thesis, a brief description is included here.

1.2.3.1 Mechanical

Mechanical theories suggest that optic nerve damage is induced by mechanical stress and strain on the posterior structures of the eye, particularly at the level of the lamina cribrosa (Quigley 1985; Weinreb et al. 2014). An increased resistance to aqueous outflow through the trabecular meshwork occurs in those with POAG, leading to an elevated IOP (Weinreb et al. 2014). Studies modelling the optic nerve head as a biomechanical structure have suggested that the connective tissues of the optic nerve head are constantly subject to IOP-related stresses, which increase with elevation of IOP; mechanical stresses are deemed physiological or pathophysiological depending on the response of the optic nerve head tissues (Bellezza et al. 2000; Burgoyne et al. 2005). Pathophysiological stresses may cause mechanical failure within the lamina cribrosa, scleral canal wall, and peripapillary sclera, leading to the classically described posterior displacement and excavation of the optic nerve head (Quigley 1985; Burgoyne et al. 2005). There have been reports of compression of the lamina cribrosa plates in glaucoma, and one such theory suggests this could cause a direct, mechanical compression of the ganglion cell axons, or affect them indirectly by compromising blood flow (Fechtner and Weinreb 1994).

As previously stated however, glaucoma can develop in the absence of a raised IOP, and there have been suggestions of an increased susceptibility to IOP with age. Crish et al. (2010), investigating transport deficits along the visual pathway in mice with glaucoma, found that a given IOP was more likely to induce transport loss in the superior colliculus of older animals. Armaly (1969) also observed that, when following a group of normal subjects, the individuals who subsequently developed glaucoma also had systemic conditions, in addition to a raised IOP.

As it is well established that the majority of those diagnosed with OHT do not appear to progress to glaucoma (Perkins 1973; Kass et al. 2002), coupled with the knowledge that NTG is often clinically indistinguishable from glaucoma associated with an elevated IOP, with respect to optic nerve head appearance and visual field defects (Anderson 2003), we must conclude that there are other factors at play.

1.2.3.2 Vasogenic

Vasogenic theories suggest that damage may be caused by compromise of the microvasculature at the optic nerve head (Fechtner and Weinreb 1994). Harrington (1964) hypothesised that glaucoma was caused by a changing ratio of ophthalmic artery pressure to IOP, causing the blood flow to decrease in the optic nerve. In the case of narrowing arteries due to disease, blood flow would decrease, and as such a smaller increase in IOP could result in a significant effect. Hayreh et al. (1970) found that choroidal circulation was also implicated, and that the state of circulation at the optic disc was determined by the balance between IOP and arterial blood pressure in the choroidal vessels, peripapillary choroid, and prelaminar regions of the optic nerve head. They determined that a reduction in blood pressure at the optic nerve head had a similar effect to an increase in IOP, and defined glaucoma as a disease in which the normal balance between these regions was disturbed. More recent evidence has suggested that the balance between diurnal and nocturnal blood pressure may also have an impact on glaucoma, as a > 20% nocturnal reduction in blood pressure correlated with a greater severity of visual field damage (Pillunat et al. 2015).

Studies modelling the optic nerve head as a biomechanical structure have theorised that the previously described mechanical stresses result in a combination of ischaemia,

due to IOP-related occlusion of laminar capillaries and/or a reduction in nutrient diffusion, and physical compression of ganglion cell axons, as the connective tissues are subject to strain within their elastic limits; once beyond the elastic limits, continued nerve fibre damage could occur at lower IOP levels (Burgoyne et al. 2005).

There have also been several genes associated with glaucoma, however, these appear to account for a minority (10%) of all cases (Weinreb et al. 2014).

1.2.4   Site of damage

Although the location of damage is widely understood to be the axons of the retinal ganglion cells at the level of the lamina cribrosa, there is evidence to suggest that subsequent damage also occurs at other locations along the visual pathway. Yücel et al. (2000) identified a reduction in both the number of neurons and the volume of magnocellular and parvocellular layers in the lateral geniculate nucleus (LGN). Crawford et al. (2000) identified a reduction in cytochrome oxidase reactivity (a measure of metabolism, used as an indicator of the visual afference from retinal ganglion cells) in the magnocellular and, to a greater extent, parvocellular layers of the LGN, and followed this through to a reduction in input to layer 4C in V1 of the visual cortex. Both these studies were conducted on monkeys with experimental glaucoma.

A more recent study suggests that earlier indications of glaucoma may, in fact, be found in the brain, at the level of the superior colliculus, rather than at the retina (Crish et al. 2010). This study was carried out on mice and it was found that, as in Alzheimer's and Parkinson's disease, there was a 'distal-to-proximal' progression, i.e. neuronal stress manifested early as deficits in axonal transport at distal sites (superior colliculus), before affecting proximal sites (retinal ganglion cells).

1.2.5   Identification of glaucoma

The identification of POAG is a challenge, and the three primary methods used are the measurement of IOP, examination of the optic nerve head, and functional measurements of the vision via perimetry.

IOP measurement alone, as already alluded to in section *1.2.2*, cannot be relied upon for a diagnosis of glaucoma, as some individuals will exhibit a normal IOP in the

presence of glaucomatous damage, and others will exhibit an elevated IOP in the absence of glaucomatous damage. The accepted limits of IOP normality are based on the distribution of IOP in the population. A review of IOP studies by Colton and Ederer (1980) found that a right-skewed IOP distribution was consistently identified. Leydhecker et al. (1958; 1959) first explained this as two overlapping distribution curves: the main collective, with a Gaussian distribution, indicates the true 'normal' range of IOP, and the partial-collective of the right-skew (> 21 mmHg) indicates the presence of glaucoma (Davanger and Holter 1965; Colton and Ederer 1980). However, these statistical calculations of the normative limits of IOP are based on the assumption of a Gaussian distribution in the absence of disease, which may not hold true.

Davanger and Holter (1965) presented a model based on effective pore diameter of the trabecular meshwork, and concluded that a right-skewed distribution of IOP was not unexpected, advocating that the limits of normality could only be determined clinically, rather than statistically. The review by Colton and Ederer (1980) further concluded that upper limits were arbitrary, stating that there was 'no clearcut upper limit of normality that can distinguish satisfactorily between normal eyes and those that have or will develop glaucoma'.

Given that IOP measurements can be of limited value, there has been much debate over whether structural changes in the optic nerve head, or functional changes in the visual field, are identifiable first in the initial stages of glaucoma development.

1.2.5.1 The structure-function relationship

While there is some discrepancy as to whether IOP differs significantly with age (Augsburger and Terry 1977; Colton and Ederer 1980; Costagliola et al. 1990), both retinal structure and function have been demonstrated to decline with age (Balazsi et al. 1984; Haas et al. 1986; Katz and Sommer 1986; Heijl et al. 1987), such that the distinction between normal and pathological decline can be difficult to differentiate.

Several studies have investigated the histology of normal and glaucomatous retinal ganglion cells, and a few have attempted to link these to available visual field data (Quigley et al. 1982; Balazsi et al. 1984; Quigley et al. 1987; Quigley et al. 1988;

Harwerth et al. 1999; Kerrigan-Baumrind et al. 2000; Swanson et al. 2004). It becomes clear from these studies that there is substantial variability in ganglion cell estimates, in both healthy and glaucomatous samples. Age was found to have a significant effect on ganglion cell axon number, with a loss of approximately 3.07% (over 50,000 axons) per decade (Balazsi et al. 1984; Kerrigan-Baumrind et al. 2000), although given the large differences between individuals noted with only a sample size of sixteen subjects between 3.5 and 82 years (Balazsi et al. 1984), this figure is unlikely to be truly representative. The large variability in retinal ganglion cell numbers, both within and between studies, may confound the distinction of normal aging from pathological disease. Balazsi et al. (1984) noted that the length of time it took for specimens to be 'fixed' post mortem had a substantial effect on the axon numbers, and Quigley et al. (1988) commented on the poor preservation of their samples (despite prompt fixing post mortem), indicating that some of this variability may be explained by the quality of the histological samples.

It is often stated that visual field results may be normal up to a loss of ~40% of retinal ganglion cells. This concept initially arose from several studies by Quigley et al. (1982; 1989; 2000), however the data presented in these studies do not fully support this statement. In Quigley et al. (1982), a single sample was reported as having 40% fewer ganglion cells than healthy controls in the absence of a visual field defect. There is uncertainty over the time at which this sample was fixed post mortem, which, as previously noted, may impact results. Quigley et al. (1982) state that all samples were fixed 'up to 24 hours post mortem', yet Balazsi et al. (1984) reported that their outliers occurred in those samples fixed more than 20 hours post mortem. If this one sample from Quigley et al. (1982) was fixed more than 20 hours post-mortem, the loss of ganglion cells due to glaucoma would be confounded by the loss of ganglion cells due to fixation time. They did attempt to account for differences due to fixation times by using control eyes with similar fixation schedules, although only five control samples were analysed, and only one of these samples had available visual field data. Additionally, not all glaucomatous samples had available visual field data, and patients from which some samples were taken had significant co-morbidity, for example toxic amblyopia and papilloedema. These factors highlight some of the difficulties

associated with histological studies, and may have confounded some of the study conclusions.

It is also worth noting that the visual field data from this study were from kinetic, manual perimetry, conducted by different examiners. As considerable variability has been identified between examiners, and automated, static perimetry has generally been acknowledged as more accurate than kinetic, manual perimetry (Lynn 1969; Trobe et al. 1980; Beck et al. 1985), it is possible this could account for the lack of field defects in subjects with substantial loss of retinal ganglion cells.

More recent studies that utilised SAP (Quigley et al. 1989; Kerrigan-Baumrind et al. 2000) corroborated the findings of Quigley et al. (1982), however Malik et al. (2012), in their critique of these studies, have highlighted several concerns with the data presented. Normative data were taken from a small number of samples which, given the substantial variability of retinal ganglion cell numbers reported between studies, may be insufficient to truly distinguish healthy from glaucomatous retinal locations, or determine the extent of damage. Despite this however, Malik et al. (2012) note that only four data points had a retinal ganglion cell count below the normal 95% confidence limits in the absence of a perimetric defect (Quigley et al. 1989). In critiquing a later study by the same group (Kerrigan-Baumrind et al. 2000), Malik et al. (2012) also note several glaucomatous eyes with an abnormal mean deviation on visual field testing, yet identified as having ≥ 100% of normal retinal ganglion cell numbers.

When considering whether structural optic nerve head changes precede functional visual loss, it is important to consider what happens to retinal ganglion cells prior to cell death. As Quigley (1985) noted, if retinal ganglion cells were subject to a period of dysfunction prior to cell death, an abnormal visual field may be detectable in a sufficiently sensitive test. As retinal ganglion cell dysfunction has been observed prior to cell death (Morgan et al. 2000), it is reasonable to expect that this dysfunction should be identifiable in a suitably sensitive functional test.

It has been demonstrated that the standard 24-2, or 30-2 test grid, in which stimuli are presented at 6° intervals, is not optimal with respect to the distribution of retinal

ganglion cells (Garway-Heath et al. 2000a). A test grid incorporating locations that are more representative of the retinal ganglion cell density, as suggested by Asaoka et al. (2012), has been shown to demonstrate a stronger correlation between structure and function.

There has been extensive debate over whether glaucoma causes diffuse or localised loss of ganglion cells. Localised loss of perimetric sensitivity may be more indicative of glaucoma than diffuse loss, as a diffuse loss of sensitivity is less specific and can occur in a wider range of conditions, although a localised visual field defect is not diagnostic of glaucoma (Harrington 1964). Asman and Heijl (1992) claimed that a diffuse defect was *not* indicative of glaucoma, although this has since been contested (Drance 1991; Chauhan et al. 1997; Henson et al. 1999; Artes et al. 2005a; Artes et al. 2010). Most studies agree that a combination of both localised and diffuse defects will be seen in glaucoma, although there is some discrepancy over whether one occurs earlier than the other (Armaly 1969; Hoyt and Newman 1972). Henson et al. (1999) noted a diffuse loss of sensitivity in 60% of their participants, and Artes et al. (2005b; 2010) found that almost all glaucomatous visual field progression comprised both focal and diffuse components, indicating that purely focal progression is rare.

Given that many aspects of perimetry (discussed in further detail in section *1.3*) are targeted at identifying focal visual field loss, such as those indicated in Figure 7 of Sample et al. (2004), this is an important consideration when evaluating the sensitivity of functional tests. Chauhan et al. (1997) pointed out that experiments analysing diffuse versus localised visual field loss are subject to a significant bias, as patients who display diffuse field loss only are less likely to be referred as a glaucoma suspect in the absence of other glaucomatous indicators, and as such are not available for analysis in these studies. It is likely, therefore, that the incidence of purely diffuse visual field loss is underestimated.

As a final point on the 'diffuse versus localised' debate, Quigley et al. (1982) noted that the methods by which ganglion cell count is subdivided per sector will impact on the figures generated. Generally, optic nerve head quadrants are referred to as per *Figure 1.1.A*. When segmented in this manner, Quigley et al. (1982) found an hourglass-

shaped atrophy in the presence of glaucoma, i.e. more ganglion cell axon loss in the superior and inferior quadrants. However, if the optic nerve head is subdivided as per *Figure 1.1.B*, which they acknowledged as an equally valid method, atrophy appeared much more uniform across all four quadrants, although they do note that the more traditional method of segmentation in *Figure 1.1.A* more closely resembles the positioning of the ganglion cell axons from various areas of the retina. Were the segmentation method of *Figure 1.1.B* more widely accepted however, a diffuse loss of retinal ganglion cells, and as such a diffuse visual field loss, may be more widely regarded as indicative of glaucoma.
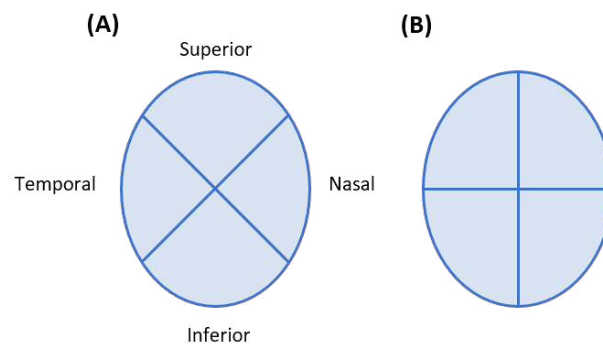
Figure 1.1 – (A) Traditional and (B) alternative methods of optic nerve head segmentation, from Figure 5 of Quigley et al. (1982).

Various studies have attempted to quantify the relationship between structure and function. Harwerth et al. (1999) indicated that the relationship between light sensitivity and ganglion cell number is curvilinear, such that a substantial loss of retinal ganglion cells must occur before functional changes are apparent, thus lending weight to the argument that structural changes occur prior to functional changes. However, Garway-Heath et al. (2000a; 2002) established that this was largely due to the plotting of a logarithmic decibel (dB) scale against the linear retinal ganglion cell numbers. They demonstrated that both pattern electroretinogram (PERG) and temporal neuroretinal rim area, plotted against perimetric sensitivity, will show a curvilinear relationship when sensitivity is displayed in dB, but a linear relationship when

sensitivity is displayed in 1/Lambert, a non-logarithmic scale, demonstrating that structural and functional changes appear to occur simultaneously. Furthermore, and as highlighted by Malik et al. (2012), Harwerth et al. (1999) do demonstrate a perimetric sensitivity loss of more than 5 dB with SAP, with a retinal ganglion cell loss of less than 10%, revealing SAP to be more sensitive to early ganglion cell loss than indicated by other studies. With the advent of sophisticated imaging techniques, it is perhaps tempting to overlook subjective visual field tests in favour of these more objective techniques, however caution should be exercised with this approach. Although objective, imaging techniques such as optical coherence tomography (OCT) are still subject to test-retest variability, and may produce both false positive (Chong and Lee 2012) and false negative results (Sayed et al. 2017). It is also important to remember that they are structural measures, and although correlations between structure and function have been identified, as yet visual perception cannot be accurately predicted from structural data (Malik et al. 2012). Objective, functional tests, such as electroretinogram (ERG), and functional Magnetic Resonance Imaging (fMRI), have been studied as a potential means of glaucoma detection (Bode et al. 2011; Jafarzadehpour et al. 2013; Preiser et al. 2013; Gerente et al. 2015; Golemez et al. 2016), but these methods are currently not widely used in a clinical setting. Additionally, the prior assertion that structural change occurs prior to functional change has been largely rejected, with a linear relationship demonstrated between the two (Garway-Heath et al. 2000b; 2002). Despite this linear relationship, it is apparent that some patients display a structural change with no apparent functional change (Sample et al. 2006), while others display a functional change in the absence of an apparent structural change (DeLeón-Ortega et al. 2006). In addition, the extent of visual field loss displayed by patients may differ substantially, despite the appearance of very similar optic nerve head structures (Malik et al. 2012), thus demonstrating some of the difficulties in quantifying the structure-function relationship. Given the many levels of processing that occur along the visual pathway, it is likely that the relationship between the cellular structures of the retina and the visual perception resulting from this processing, is indirect and convoluted, emphasising the necessity of both types of measurement in a clinical setting. Two recent reviews have highlighted the advantage of both types of measurement, indicating that neither should be

considered superior to the other (Camp and Weinreb 2017; Phu et al. 2017a). Additionally, it should be remembered that 'the purpose of treating glaucoma is not to lower pressure, or even to prevent visual field progression, but to prevent blindness' (Schulzer 1994). Glaucoma is treated to preserve quality of life, and measures of visual function thus remain vital in the understanding of patients' perception of, and interaction with, the world around them (Medeiros et al. 2015; Orta et al. 2015; Camp and Weinreb 2017). This is an important consideration which should not be forgotten in the continued pursuit for effective disease management.

1.3 Perimetry

1.3.1 Evolution of perimetry

1.3.1.1 Initial development

The value of measuring the field of vision was identified long before the term 'glaucoma' had been fully developed. There are many references to peripheral visual field evaluation dating back to 5th century B.C., when Hippocrates observed and described a hemianopsia (Johnson et al. 2011), however Albrecht von Graefe is the person credited with introducing visual field testing and perimetry to the clinical setting (von Graefe 1856). He recognised the value in mapping the visual field, and did so using a flat board divided into numbered squares. The patient fixated a central target and the examiner moved an object around the board, marking where it seemed to disappear. The marked squares were then connected with a line, indicating the extent of the patient's field of vision, i.e. as would now be referred to as isopters (Simpson and Crompton 2008a). Examples of these field plots are shown in *Figure 1.2*. Aubert and Foerster introduced the arc perimeter in 1869, in an attempt to measure a greater extent of the visual field at a constant distance from the eye (Simpson and Crompton 2008a; Johnson et al. 2011), an example of which is shown in *Figure 1.3*. There were reported difficulties in achieving a uniform background adaptation level with this, triggering the development of the bowl perimeter. This had the added advantage of eliminating background distractions.
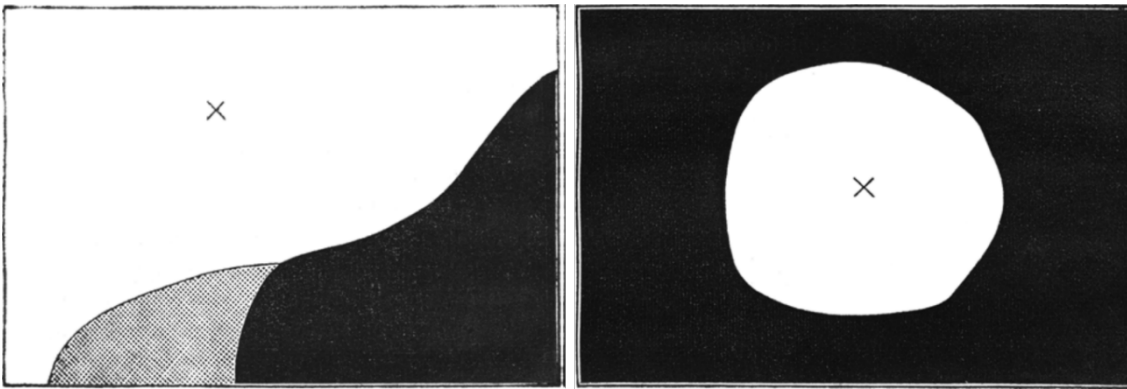
Figure 1.2 – Examples of visual field defects as mapped by von Graefe (1856).
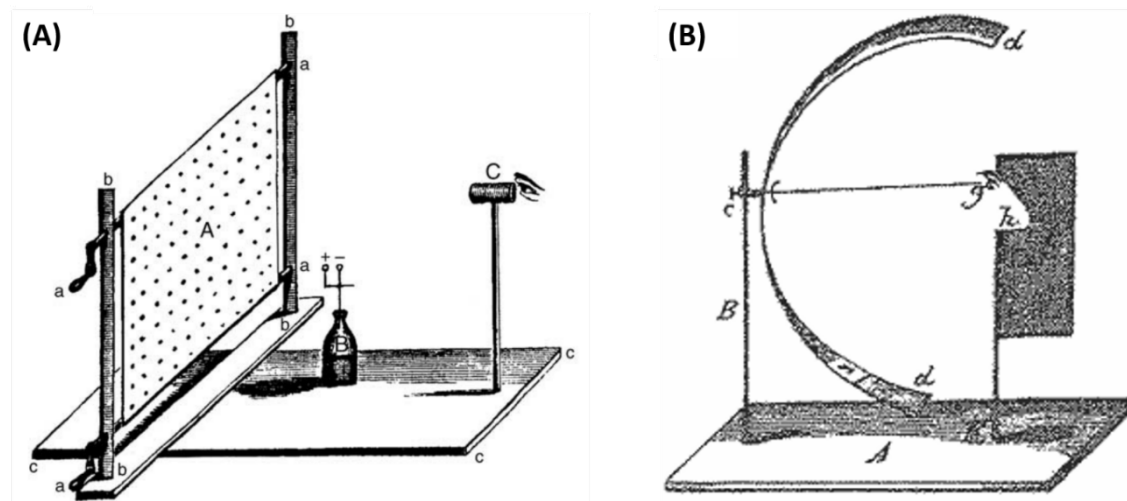
Figure 1.3 – (A) Flat, tangent screen (Aubert and Foerster 1857), developed to (B) arc perimeter (Foerster 1869; http://www.perimetry.org/index.php/measurement-of-the-visual-field-limits-the-perimeter, accessed on 06/08/17).

Two examples of bowl perimeters are shown in *Figure 1.4*. The appearance of these perimeters is surprisingly familiar, despite being developed and described almost 150 years ago. Jeaffreson (1873) provides a detailed description of his perimeter, in which a gas lamp and mirror were used to project both a fixation point, and a moving target onto a white, hemispherical background. The patient rested their chin on a chin rest to maintain a test distance of 30 cm, and two sets of diaphragms were used, one with different sized apertures and one with different colours of glass.
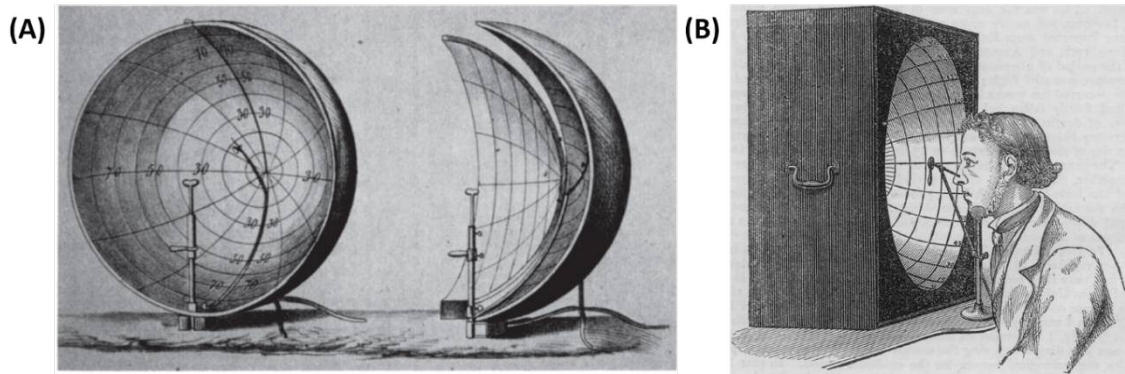
Figure 1.4 – Examples of early bowl perimeters by (A) Scherk (1872; http://www.perimetry.org/ index.php/measurement-of-the-visual-field-limits-the-perimeter, accessed on 06/08/12) and (B) Jeaffreson (1873).

From *Figure 1.4* and the descriptions provided, it is not difficult to imagine how this evolved into today's model. However, early literature reveals many differing viewpoints as to the 'correct' method by which to conduct an accurate perimetric assessment. These opinions appear to be largely due to the individual perimetrist's own experience, rather than results of controlled experiments. Large variations in fundamental parameters, such as background luminance, stimulus size and colour, and the observer's distance from the instrument were reported. In addition, the ocular and neurological conditions under assessment were not yet fully understood.

Jeaffreson (1873) noted that many hospital surgeons relied on the 'old custom' of confrontation testing, even though it 'cannot be registered, and must leave but a doubtful impression on the mind of the surgeon', whereas Elliot (1920), nearly 50 years later, advocated that a 'rough test' (i.e. confrontation) was sufficient to identify a reduction in the field of vision due to glaucoma, and that 'no apparatus was needed'.

Bjerrum reported in 1889 that he could plot visual field defects more successfully on the back of his consulting room door than with the use of a perimeter, thus leading to his development of the Bjerrum, or tangent, screen (Traquair 1927; Ehlers 1981; Simpson and Crompton 2008a; Johnson et al. 2011). Some practitioners preferred the use of this tangent screen, typically a black, flat screen with white stimuli (Traquair 1935; 1939; Ehlers 1981; Simpson and Crompton 2008a; Johnson et al. 2011), while others plotted visual fields using an arc or bowl perimeter (Jeaffreson 1873; Smith

1925; Simpson and Crompton 2008a; Simpson and Crompton 2008b; Johnson et al. 2011), which typically had a white, illuminated background.

Visual field testing with tangent screens were generally conducted at a greater test distance than with an arc or bowl perimeter (Ehlers 1981); it was much more practical to set a closer test distance for arc and bowl perimeters, with most set somewhere between 25 and 35 cm (Jeaffreson 1873; Smith 1925; Simpson and Crompton 2008a; Simpson and Crompton 2008b). Von Graefe originally set his test distance at several feet (von Graefe 1856), and Traquair (1935; 1939) advocated the use of the tangent screen at 1-2 m, claiming this was more accurate; if detailed testing of central visual field was required, he recommended an increase in test distance to three or four metres, claiming this was more 'satisfactory' than decreasing the stimulus area. However, the discussion of Traquair (1939) includes comments from another practitioner, Dr L. Peter, who gives his opposing view that visual field tests were more accurate when conducted at 33 cm, and afforded earlier identification of glaucomatous defects. Cushing, a pioneer of neurosurgery, preferred the use of an arc perimeter, with a test distance of 28.6 cm for peripheral fields, and a tangent screen with a longer working distance of 2 m for central fields (Simpson and Crompton 2008b).

Differing views regarding room conditions were also apparent, with some preferring a dark room and others preferring ambient room lighting (Simpson and Crompton 2008b), although Traquair declared that it made no substantial difference in the detection of visual field defects (Traquair 1939). The peripheral extent to which visual fields were measured also differed between practitioners, e.g. Smith (1925) examined out to 40° or 50°, whereas Traquair (1939) did not feel the need to go beyond 26°.

The parameters of the stimulus itself also differed between practitioners. Bjerrum and Rönne used differing stimulus areas (Johnson et al. 2011), and Traquair declared that reducing the area of the stimulus was more effective than reducing the intensity or altering the colour (Traquair 1935; 1939). Some practitioners advocated that routine colour perimetry was more accurate than a white stimulus, whereas others did not feel a coloured stimulus revealed anything different to that of a white one (Traquair 1939).

Jeaffreson's perimeter (1873) had the ability to alter the area of the stimulus and seems to indicate that this was the primary method of measuring visual thresholds, although it also included two methods of presenting colour stimuli.

1.3.1.2 Standardisation of perimetry

Given the personal preferences amongst practitioners as to the optimum method of visual field testing at that time, there was likely substantial confusion with respect to the analysis of visual field plots and what they revealed. Thomasson (1934), in fact, wrote a paper entitled, 'A plea for greater uniformity in methods of field taking', highlighting the problems associated with the lack of standardisation. He acknowledged that exact standardisation would be difficult, likely as it was unclear which method was genuinely superior, identifying that many variables could affect the result of a visual field plot. His great concern was the inattention given to the conditions under which visual field tests were being conducted, and that many results were being published without disclosure of the exact conditions under which they were obtained. His paper identifies, in particular, the area and colour of the test stimulus, the instrument used, illumination conditions, and the charts on which thresholds were recorded.

Ferree and Rand (1920a; 1920b) were commissioned to attempt a standardisation of perimetric illumination. Like Thomasson, they commented that a lack of precision and standardisation of test parameters would inhibit the production of reproducible results, rendering measurements almost meaningless. In particular, they emphasised stimulus intensity, fixation accuracy, and uniformity of both the room illumination, and the illumination of the instrument (in this study, an arc perimeter), as those most crucial to control in a clinical setting. They examined and made recommendations on the different methods of ensuring a consistently illuminated stimulus at each test location, and recommended the use of a background the same colour as the stimulus (i.e. white-on-white), such that the stimulus would 'disappear completely' when sub-threshold.

Their attempts to steady fixation were logical, but are no longer in use. They advocated the use of a small circular mirror as a fixation target; if the observer could

see their own eye in the mirror then they knew they were in the correct position. A measuring rod was used to confirm that the observer was at the correct distance from the instrument, and the use of a bite-bar was deemed the simplest and most effective way to guard against any movement of the observer's head.

The work of Ferree and Rand (1920a; 1920b) unfortunately did not achieve the desired perimetric standardisation at that time. Goldmann (1945b; 1999) commented that their work was being 'disregarded to an astonishing degree', however by building on the studies of Ferree and Rand (1920a; 1920b), Goldmann did manage to achieve a perimetric standard, where Ferree and Rand had not. Originally developed in 1945, the Goldmann Perimeter is credited as being probably the most important contribution to modern perimetry (Johnson et al. 2011), and came to be considered the reference standard in visual field measurement. The instrument was a bowl perimeter of uniform background illumination with a moving projection system, and was able to perform both kinetic and static perimetry. Stimuli could be altered in area, intensity and colour, as required by the perimetrist (Johnson et al. 2011).

Goldmann felt that 'light sense' perimetry (which had been mainly white-on-black up to that point, while he opted for white-on-white) had a greater following than colour perimetry, and so he elected to concentrate on this (Goldmann 1945a; 1999). As he felt colour perimetry could be helpful when defects were present, however, this capability was also included. He determined that the optimum speed at which to perform kinetic perimetry, thus obtaining the largest field and least variability of results, was approximately five degrees per second. He recognised that the ratio of background luminance to stimulus intensity, $\Delta I/I$, was important, and that small changes in this ratio could affect the size of the visual field. He did a lot of work on the relationship between stimulus area and luminance, which is discussed in more detail in section *1.3.2.3*.

1.3.1.3 Automation of perimetry

The standardisation of perimetry, in the form of the Goldmann Perimeter, did not eliminate all sources of variability. Studies identified many potential sources of variability, some still related to instrumental hardware, and others human. Much

research was now focused on the human variability and possible methods of reducing/eliminating this.

It was recognised, and extensively reported upon, that both the examiner and the patient were considerable sources of error and variability in visual field testing (Lynn 1969; Heijl and Krakau 1975; Fankhauser et al. 1977; Portney and Krohn 1978; Trobe et al. 1980; Beck et al. 1985). While patient-related variability was more difficult to reduce, many studies concluded that the next logical step was to develop a completely automated perimeter, believing this would remove all examiner-related sources of error. Perhaps Heijl (1976) put it best when he wrote 'We find it likely that instead of acquiring increased skill, the perimetrist runs the risk of getting bored from dealing with perimetry day in and day out, unless he/she has got the psychic constitution of an automatic machine'. The belief was held that fully automated perimetry would result in quicker, more efficient testing, with more accurate, repeatable results.

Most initial work was concerned with attempting to create an automated version of the Goldmann Perimeter, but there were many technical difficulties associated with the kinetic aspects of this (Fankhauser et al. 1977; Portney and Krohn 1978; Gloor 2009), and any resulting programs were inferior to manual, kinetic perimetry (Portney and Krohn 1978). This was the main contributing factor that led to the automation of *static* perimetry. Some work was based on the older tangent screens (Portney and Krohn 1978), but it was the Octopus Perimeter, based on the Goldmann Perimeter and created by Fankhauser and his team, that is credited as being the first truly automated perimeter, an example of which is shown in *Figure 1.5* (Fankhauser et al. 1977; Gloor 2009). Although the original brand of automated perimeter, it has been overtaken by the Humphrey Field Analyzer series as the primary automated perimeter (Gloor 2009), despite substantial similarities between them. The Humphrey Field Analyzer is now considered to be the reference standard of perimetry, in accordance with the National Institute for Health and Care Excellence (NICE) guidelines (NICE 2009; 2017).

Figure 1.5 – The original Octopus Perimeter (Fankhauser et al. 1977).

Many studies report that the new automated, static perimeters were more accurate than the manual, kinetic perimeters, however given the examiner-induced variability associated with manual perimetry, it is difficult to accurately determine whether this was due to the standardisation achieved by the automation, or whether static stimuli were superior to kinetic stimuli (Lynn 1969; Trobe et al. 1980; Beck et al. 1985). In addition, static and kinetic techniques involve quite different stimulus parameters; a static strategy incorporates a Goldmann III stimulus (area 0.15 deg^2), which varies in luminance, while a kinetic strategy often incorporates a Goldmann I stimulus (0.01 deg^2) of fixed luminance. As such, it can be difficult to make direct comparisons between them, especially in a statistically meaningful way. It is also likely that practitioners were more familiar with the kinetic, manual methods of visual field testing, and the resulting isopter plots, than with the newer, static, automated methods and the resulting output. For example, Beck et al. (1985) commented that they could not quantify the differences in the field plots, and therefore compared manual and automated field plots by eye alone, using only the grayscale plot of the Humphrey Field Analyzer. As the grayscale plot is not considered to be a very sensitive representation of the measured visual field (Heijl et al. 2012), this may introduce a level of bias to the conclusions drawn by the study.

More recently, there have been successful attempts to automate kinetic perimetry, and a program is now incorporated into the Octopus 900 perimeter (Haag-Streit 2014), and the Humphrey Field Analyzer 3 (Zeiss 2015). It is identified as a 'semi-automated' program, as examiner involvement is still required, and as such may still be subject to some examiner-related variability. Several studies have attempted to compare these semi-automated kinetic programs with established static programs, although many with a view to examining neurological field defects, an area in which the manual Goldmann Perimeter is persistently used, rather than for use in glaucoma management. Rowe et al. (2013) compared the kinetic program on the Octopus 900 with the FF-120 program on the Humphrey Field Analyzer, in which the stimulus luminance remains constant, in keeping with that of the kinetic strategy. They determined that kinetic testing was of significantly shorter duration than the static testing, and established an 80% agreement between the kinetic program on the Octopus 900, using the I4e stimulus, and the FF-120 program on the Humphrey Field Analyzer. It is worth noting that test duration was analysed with an unpaired t-test; given that all participants undertook both testing strategies, it would have been more appropriate to use a paired t-test, although given that mean duration was 4.54 ± 0.18 minutes for the kinetic strategy, and 6.16 ± 0.12 minutes for the static strategy, it is likely this would still achieve statistical significance. It is perhaps not a direct comparison between testing strategies, given the differences in stimulus area and luminance between tests, in addition to the differing instruments used for the two testing strategies. It may be more appropriate to equate stimulus area and luminance, the extent of the visual field being measured, and to conduct tests on one instrument alone, to ensure it is the testing strategies that are being directly compared.

More specific to glaucoma, Nowomiejska et al. (2015) and Mönter et al. (2017) have investigated the use of semi-automated kinetic perimetry in addition to SAP, rather than as a replacement, and particularly its use in testing further peripheral locations than the standard central 30°. Both these studies used a Goldmann III stimulus with the kinetic program, and conducted both kinetic and static tests on the same instrument (Octopus 900), to afford a more direct comparison between testing strategies. Nowomiejska et al. (2015) noted that additional information could be

obtained in end-stage glaucoma (mean deviation (MD) < -20 dB) with kinetic perimetry in comparison to SAP, both with the standard testing parameters of the central field, and further into the periphery, and Mönter et al. (2017) noted that patients with similar central visual fields may exhibit very different perimetric sensitivities further into the peripheral field. Both studies concluded that the examination of further peripheral locations more readily afforded by the kinetic strategy may provide additional information in the investigation of glaucomatous visual field defects.

Despite the obvious advantages of automating perimetry, it quickly became clear that automation had not eliminated variability completely (Wilensky and Joondeph 1984). This is discussed in more depth in section *1.3.3*, although prior to this discussion, it is prudent to examine the parameters incorporated in automated perimeters.

1.3.2 Perimetric parameters

1.3.2.1 Background luminance

It is perhaps unfortunate for this author in the pursuit of perimetric understanding that much work, both in the development of the Goldmann Perimeter and in later developments of automation, was reported largely in French and German (Goldmann 1945b; Goldmann 1946; Dubois-Poulsen et al. 1952; Gafner and Goldmann 1955; Fankhauser and Schmidt 1958; Fankhauser and Schmidt 1960), little of which has been translated to English. As such, the rationale for the incorporation of various perimetric parameters, including background luminance, extent of field tested, stimulus duration, and stimulus area, may be unclear.

Many of the parameters Goldmann selected for his perimeter have endured through to more recent perimeters, and are still in use in current designs. It appears that several parameters may have been imported to automated designs from the manual Goldmann Perimeter to aid with interpretation of subsequent measurements (Heijl 1984). For example, the uniform background luminance of 10 cd/m^2 was used in the Goldmann Perimeter, and is incorporated in some of the most current models of automated perimeter, such as the Humphrey Field Analyzer (Zeiss 2014; Zeiss 2015). The original Octopus Perimeter had a background luminance of 1.3 cd/m^2, in an attempt to increase the dynamic range of the instrument (Fankhauser 1979), but more

recent models have the option of a higher background luminance, at 9.9 cd/m^2, similar to the Goldmann Perimeter (Haag-Streit 2014). Anderson (2006) speculates that Goldmann chose this background luminance as it places healthy observers within the range where ΔI/I is constant, i.e. where Weber's law operates (Glezer 1965), although Fankhauser (1979) contested this, stating that validity of Weber's law could only be expected if the background luminance was at least two-to-three times larger. Anderson (2006) also noted that Goldmann was limited by the artificial illumination levels available at the time, with 10 cd/m^2 being the highest available while still providing an adequate dynamic range. Heijl (1984) advocated that the 'standard Goldmann background' requires less pre-adaptation than lower background levels. He stated it was necessary for earlier automated perimeters to operate at lower background levels to increase the dynamic range of the instrument, due to limitations in the output of the light source used, however there was no clinical advantage with this lower background luminance, and 10 cd/m^2 was recommended in future instruments.

Crosswell et al. (1991) investigated two low photopic levels of 10 and 1 cd/m^2, to mimic the background intensities incorporated by the frequently used Humphrey Field Analyzer (Zeiss 2014; Zeiss 2015) and the Octopus Perimeter (Haag-Streit 2014), and measured these against two mesopic levels of 0.1 and 0.01 cd/m^2. They reported that long-term homogeneous fluctuations were lower at 10 cd/m^2, but that long-term heterogeneous and interindividual fluctuations were lower at 1 cd/m^2. Total fluctuation was similar between both background intensities, and they concluded that they should give comparable results. They found that the mesopic intensities of 0.1 and 0.01 cd/m^2 increased all types of fluctuation except long-term homogeneous fluctuations, concluding that these levels would create more uncertainty if used in perimetry. Scotopic intensities were not tested, as the authors acknowledged that this would be clinically impractical, given the longer periods of adaptation, coupled with the difficulties of mobility for patients in a completely darkened room. It would perhaps have been beneficial to test higher photopic intensities as well, although this would reduce the dynamic range of any instrument.

1.3.2.2 Test distance and extent of field tested

Testing distance is generally set at 30 cm, as per the Goldmann Perimeter (Haag-Streit 2014; Zeiss 2014; Zeiss 2015; Haag-Streit AG n.d.). As previously discussed in section *1.3.1.1*, there was some debate amongst practitioners as to whether a closer or further working distance gave a more accurate result, but as Goldmann chose a bowl perimeter for his design, this would have limited the working distance options.

In SAP, the visual field is often only tested up to 30° from fixation, differing from the Goldmann Perimeter. Manual, kinetic perimetry generally tested a greater extent of the peripheral visual field, as isopters were constructed by moving a stimulus from an 'unseen' to a 'seen' eccentricity. It was noted that the contraction of peripheral isopters measured from manual, kinetic perimetry may not be specific to glaucoma, and other possible causes had to be ruled out before a diagnosis of glaucoma was confirmed; Armaly (1969) designated this a 'diagnosis by elimination', and it was thought that if the defect involved the central field, this was a better indication of glaucoma. The routine testing of the far periphery was debated amongst practitioners, with Blum et al. (1959) stating that measurement beyond the central 30° was not necessary, and Fankhauser (1979) emphasising the important information that could be detected in the far periphery.

Of course, with static perimetry, less of the visual field can be covered within a reasonable timespan. As Portney and Krohn (1978) acknowledged, to test all 360° of each sensitivity circle, separated by 1°, would require 10,800 retinal locations within the central 30° alone. As this is evidently clinically impractical, it may be that with the advent of static perimetry, testing of the central 30°, rather than a full peripheral examination, was deemed an acceptable compromise.

1.3.2.3 Stimulus

Duration

The visual system integrates a light stimulus over time, such that, over a range of durations, there is a reciprocal relationship between stimulus duration and threshold luminance; a static stimulus presented for 0.002 seconds will be approximately twice as visible as one presented for 0.001 seconds, known as Bloch's law of temporal

summation (Bloch 1885). However, beyond a specific exposure time, temporal summation is incomplete, and further exposure of the stimulus will not result in a higher level of visibility. Summation is reported as largely complete by 0.1 seconds, although additional summation may occur up to one second under specific circumstances (Bruder and Kietzman 1973; Anderson and Patella 1999).

Automated perimeters utilise temporal summation by generally employing stimulus durations of 0.1 or 0.2 seconds (Haag-Streit 2014; Zeiss 2014; Zeiss 2015), such that small variations in shutter duration have little effect. A stimulus duration of 0.2 seconds or less is also said to stabilise observer fixation, as it does not allow sufficient time to take up fixation of the stimulus (Anderson and Patella 1999).

The initial work involved in the development of the Humphrey Field Analyzer used a stimulus duration of 0.5 seconds (Heijl and Krakau 1975). Later analysis of the variability of stimulus durations between 0.065 and 0.5 seconds (Pennebaker et al. 1992), found that there did not appear to be an obvious trend in threshold fluctuation, with the exception of some increase in variability in the inferior visual field with a duration of 0.5 seconds. It should be noted that this experiment only involved healthy observers, such that the effects of stimulus duration on those with a visual field defect were unknown at the time.

A more recent study on critical duration, i.e. the maximum duration at which Bloch's law holds, has reported a substantially shorter critical duration than the previously reported 0.1 seconds for a Goldmann III stimulus (Mulholland et al. 2015a). The same study also reported that it was unlikely that contrast thresholds were independent of stimulus duration at the standard 0.1 or 0.2 seconds. Additionally, a further study by the same group (Mulholland et al. 2015b), found a different critical duration in glaucoma compared with healthy observers. As such, they recommended a shorter stimulus duration, estimating that the disease signal in glaucoma could be boosted by approximately 200%, if a stimulus of equivalent duration to that of the critical duration in healthy observers were utilised.

Area and luminance

An achromatic, circular stimulus of area 0.15 deg^2, i.e. a Goldmann III stimulus, is commonly used in SAP. Goldmann 0 to V are standard perimetric stimuli, based on Goldmann's original work in 1945, and modified only slightly since (Goldmann 1945a; Goldmann 1999). He attempted to find the correlation between stimulus area and luminance, i.e. the change in area and the change in luminance that would produce equivocal values. Using kinetic perimetry, his data indicated that between Goldmann I and II, there was partial summation, and that a fourfold increase in area was approximately equivalent to an increase in intensity by a factor of 3.16. Goldmann calculated from this a summation coefficient of approximately 0.8. He found that this coefficient was lower with larger stimuli, and also noted that it differed with eccentricity from fixation, however concluded that for the purposes of perimetry, a constant value of 0.8 could be assumed.

Goldmann designed his six stimuli with areas that differ by 0.6 log units, based on the ratio of 1:4, ranging from 0.002 deg^2 to 2.32 deg^2, whereas the original four standard luminance values differed by 0.5 log units. Therefore, in regions where this relationship between area and luminance holds, an equivalent isopter could be plotted by changing stimulus area or stimulus luminance accordingly. Goldmann 0, the smallest of the six, has largely fallen out of use.

When automated perimetry was initially developed, the Goldmann III was deemed the most suitable stimulus, although the reasons for this decision are unclear. Fankhauser (1979) and Heijl (1984) state that a 0.15 deg^2 stimulus increased the dynamic range, and was more resistant to the effects of optical blur, than the 0.01 deg^2 (Goldmann I) stimulus that was commonly used at the time in kinetic, manual perimetry. Swanson (2013), however, refers to the 'historic accident' that led to its usage, and declared that this choice of stimulus area was made by the engineers designing the original automated perimeter, and was not a scientific decision. Maximum luminance in automated perimetry was initially set at 318.3 cd/m^2, in keeping with that of the Goldmann Perimeter, although it has since been extended to increase the dynamic range of the instrument, and is currently set at 3183.1 cd/m^2 in the Octopus 900 (Haag-Streit 2014) and Humphrey Field Analyzers (Zeiss 2014; Zeiss 2015).

There has been much work conducted since 1945 on the most appropriate stimulus area to use in perimetry, although much of the research has remained within the confines of the established Goldmann stimuli, rather than exploring other alternatives (Wilensky et al. 1986; Choplin et al. 1990; Gilpin et al. 1990; Wall et al. 1997; Gardiner et al. 2015); the references cited here are a very small selection of those available.

Wall et al. (1997; 2013) have advocated the use of a larger target, Goldmann V or the extrapolated Goldmann VI, as they determined that variability of results was lower with larger targets. This is discussed in more detail in chapter three, but briefly, it is of note that differing stimulus areas can be difficult to compare accurately in this manner. Consider a Goldmann III stimulus of 30 dB (3.18 cd/m^2); if the overall light energy of the stimulus is determined according to *Equation 1.1*, this stimulus would have an energy value of 0.09 cd/m^2.s.deg^2. With a change of 1 dB, to 29 dB (4.01 cd/m^2), this stimulus now has an energy value of 0.12 cd/m^2.s.deg^2, a difference of 0.03 cd/m^2.s.deg^2. A Goldmann V stimulus, changing from 30 dB to 29 dB as before, would have a difference of 0.38 cd/m^2.s.deg^2, a much greater difference; this difference will increase and decrease with lower and higher sensitivities respectively, due to the logarithmic nature of the dB scale.

$$E = L \times A \times D$$

<div align="right">*Equation 1.1*</div>

$E = Energy\ (cd/m^2.s.deg^2)$

$L = Luminance\ (cd/m^2)$

$A = Area\ (deg^2)$

$D = Duration\ (s)$

The use of a Goldmann V stimulus has been found to extend the reliable sensitivity range that can be tested, thus extending the dynamic range of the perimeter (Wilensky et al. 1986; Gardiner et al. 2015), however it has been shown to be less sensitive than

the smaller Goldmann III stimulus (Wall et al. 1997; Phu et al. 2017b). As such, it may be of use, not in the identification of early glaucomatous field loss, but in more advanced disease, in which it is not possible to establish a sensitivity value with a Goldmann III stimulus.

In the original Goldmann Perimeter, if the chosen stimulus area could not be perceived, a larger area could be selected (Haag-Streit AG n.d.). Other perimeters have also incorporated differing stimulus areas, e.g. the Friedmann Analyser, in which stimulus area was increased with eccentricity (Friedmann 1966) in an attempt to account for the decreased sensitivity in the peripheral visual field. These stimulus areas were not limited to those established by Goldmann, however concerns were raised that the stimulus areas had been incorrectly determined due to a lack of near refractive correction in the preliminary studies, and that the apertures incorporated in the instrument itself to display the stimuli were not always accurately sized (Greve 1973; Henson et al. 1984). Wall et al. (2004) investigated 'Luminance Size Threshold Perimetry', incorporating a stimulus of fixed luminance which varied in area to determine threshold, all presented on a flat display screen. More recently, Hirasawa et al. (2016) investigated what they termed 'size-modulation standard automated perimetry', in which stimulus area increases once the maximum luminance of the instrument has been reached, and Kalloniatis and Khuu (2016) investigated a strategy whereby stimulus area differed according to eccentricity from fixation. Both of these strategies incorporated only those stimulus areas determined by Goldmann.

Choplin et al. (1990) noted that the database of normative values that is included in standard automated perimeters are all calculated with a Goldmann III stimulus. He attempted to calculate a 'correcting factor' to convert these normal values from a Goldmann III to a Goldmann V stimulus, although did not take into account that this may not be a constant value at different eccentricities. Additionally, Gilpin (1990) measured the variability with each of the Goldmann stimuli and concluded that both smaller and larger stimuli than Goldmann III increased the amount of fluctuation, be it short term or longer term. Given that a Goldmann III stimulus falls within the area of complete spatial summation at some retinal locations, but not others (as discussed below), a simple conversion from one Goldmann stimulus to another may not be

possible, and a more accurate method would be re-collection of normative data with healthy observers, using a Goldmann V stimulus.

Spatial summation is discussed in greater detail in section *1.4*, but is described briefly here. It describes the means by which the visual system integrates the amount of light energy over a given stimulus area (Redmond and Anderson 2011). This summing of information is fundamental to the way in which the visual system processes information.

Ricco's law states that, for a range of small stimuli, threshold intensity is inversely proportional to the area of the stimulus at threshold. The stimulus is said to undergo 'complete spatial summation', and Ricco's area is the largest stimulus area for which this law holds (Ricco 1877). Beyond this area, summation is only partial, also referred to as 'incomplete' or 'probability' summation.

Whether spatial summation is partial or complete is dependent on eccentricity from fixation, and there are large variations between individuals (Sloan 1961; Hallett 1963; Wilson 1970). Wilson (1970) calculated the maximum area of complete spatial summation at various locations of the retina, and found an average maximum area of 8.65' at 5° eccentricity, and 44' at 50° eccentricity. To equate this to Goldmann stimuli, 8.65' is between Goldmann I and II, and 44' is between Goldmann III and IV. As noted previously, Goldmann himself found these differences with eccentricity, but did not take this into account when calculating his summation coefficient and selecting his subsequent stimulus areas (Goldmann 1945a; Goldmann 1999). It should be noted that Wilson (1970) conducted his experiments with a much higher background luminance than that of the Goldmann Perimeter, which will result in a reduction in Ricco's area measurements (Glezer 1965). The findings of Sloan (1961) support this. She conducted her experiment with Goldmann 0 to V using the Goldmann Perimeter, incorporating a lower background luminance than that of Wilson (1970), and found that Goldmann II was within the area of complete spatial summation, even at the fovea. Sloan (1961) also reported that Goldmann III was within the area of complete spatial summation, although observation of the summation curves provided perhaps contradicts this statement.

Goldmann (1945b; 1999) used kinetic stimuli to find his values, whereas Sloan (1961) and Wilson (1970) used static perimetry; these differences in study design may also contribute to the differing conclusions as to whether Goldmann stimuli were undergoing complete or partial summation. The duration of the stimulus has also been found to influence measurements of Ricco's area (Barlow 1957; Barlow 1958; Davila and Geisler 1991), and Sloan (1961) has not reported the stimulus duration used in her experiments. Wilson (1970) used a duration of 0.13 seconds, consistent with the standard 0.1 to 0.2 seconds commonly incorporated in current SAP (Haag-Streit 2014; Zeiss 2014; Zeiss 2015).

Given that the Goldmann stimuli were calculated in tandem with specific luminance values, designed to give equivalent isopter values with kinetic perimetry, and given what is now known regarding spatial summation and how this relates to the established Goldmann stimuli, there is little to recommend the persistent use of these restricted area measurements in static perimetry. It appears that tradition is being continued for tradition's sake, an action which must be discouraged in the absence of scientific justification. Swanson (2013) advocated the need to investigate the parameter of stimulus area further, declaring there was 'nothing special' about the use of a single stimulus area, and that in the continued development of perimetry, exploration of stimulus areas beyond those described by Goldmann (1945b; 1999) is essential. It is possible that the continued use of various perimetric parameters, in the absence of scientific justification, may have contributed to some of the known limitations in SAP, discussed in section *1.3.3*.

1.3.3   Limitations of SAP

These limitations can be summarised into three main categories:

1. Early glaucomatous damage at a visual field location may not reliably be distinguished from normal (Wilensky and Joondeph 1984; Artes et al. 2002a; Tafreshi et al. 2009).

2. There is substantial variability in perimetric sensitivity measurements in the presence of early disease. This increases with increasing damage, to the extent where test-retest variability may range from normal sensitivity to absolute

defect; sensitivity values below ~15 dB have therefore been determined as 'unreliable'. This can impede identification of disease progression from stable disease, potentially delaying necessary treatment modifications (Wilensky and Joondeph 1984; Heijl et al. 1989a; Wall et al. 1996; Artes et al. 2002a; Russell et al. 2012; Gardiner et al. 2014).

3. There is a short dynamic range, such that advanced disease cannot be measured fully. This is partially due to the extensive test-retest variability, and partially due to a floor effect, with perimetric sensitivity determined as '0 dB' in an area with still-present vision (Wilensky et al. 1986; Artes et al. 2002a; Gardiner et al. 2014).

It is important to highlight at this stage that the term 'sensitivity', as used when referring to perimetric measurements, is somewhat of a misnomer. 'Sensitivity' refers to the ability of the cells in the visual system to detect and respond to a stimulus. This ability does not change in the short-term. However, the 'threshold', i.e. the measurement of that sensitivity, may fluctuate. The threshold is a measure of the minimum amount of stimulus light energy required to be perceived by an observer, and increases in the presence of damage. Perimetry is an attempt to measure the threshold of each test location, however the threshold reciprocal is often used as the measurement scale (dB), such that the measurement decreases in the presence of damage. This is often designated the 'sensitivity' of the test location. For the purposes of this thesis, the term 'sensitivity' will be used in this context, to mean 'perimetric sensitivity' i.e. the reciprocal of the threshold measurement.

The three limitations of SAP are illustrated in a study by Artes et al. (2002a), who investigated the repeatability of Full Threshold, SITA Standard, and SITA Fast strategies, using the Humphrey Field Analyzer. Participants completed each of the three testing strategies on four separate visits. Retest sensitivity values were plotted against baseline sensitivity values, a schematic of which is shown in *Figure 1.6*. With all three testing strategies, a substantial test-retest variability was demonstrated at test locations with a high sensitivity; the example illustrated in *Figure 1.6* is for a baseline sensitivity of 32 dB, with which retest sensitivities ranged between 24 and 35 dB, such

that healthy test locations could not accurately be distinguished from those locations with early disease. As sensitivity decreased, test-retest variability increased, until retest sensitivity values spanned a range between near-normal sensitivity and absolute defect; the example illustrated in *Figure 1.6* is for a baseline sensitivity of 4 dB, with which retest sensitivities ranged from 0 dB to 24 dB. The floor effect is observed from a baseline sensitivity of 12 dB and lower in *Figure 1.6*, in which baseline and retest sensitivities may be measured as 0 dB, but vision is still apparent at these locations as subsequent measurements are higher. It should be noted that the quoted values are estimated from Artes et al. (2002a), but are not exact values.
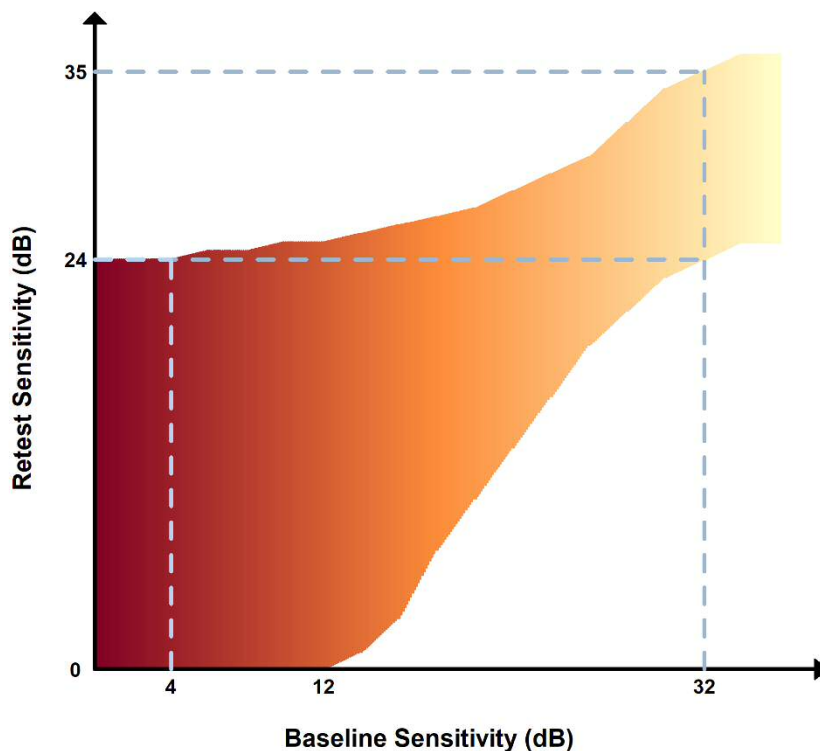
Figure 1.6 – Schematic illustrating the findings of Artes et al. (2002a). Test re-test variability increases with decreasing sensitivity.

Although sensitivity values were not exactly the same, the overall trend, as described above, was observed with all three testing strategies. This indicates that the observed test-retest characteristics are not due to the thresholding algorithm used, but to something more fundamental about SAP itself.

Decreasing sensitivity, such as that observed by Artes et al. (2002a), may be due to several causes. Heijl et al. (1987; 1989a) found that a decreasing sensitivity with increasing eccentricity from fixation, was accompanied by an increase in the variability of results obtained in healthy observers. This was true of test-retest variability, intra-test variability, and variability between subjects. In comparison, they found that eccentricity had little effect on the variability of results in glaucoma patients, although they did note an increase in variability with depth of defect.

1.3.3.1 Variability

Figure 1.6 demonstrates that the limitations of SAP may be largely due to variability in results obtained. This variability in the glaucomatous eye has been observed for almost 100 years (Ferree and Rand 1920a), and, as Artes et al. (2005b) commented, 'can both mask and mimic glaucomatous change', creating uncertainty when interpreting the resulting visual field plots. Analysis of SAP plots have suggested that four to six consecutive visual field tests are required to distinguish progression from stability (Schulzer 1994), and it has been demonstrated that conducting three tests per year identifies progression sooner than if less are performed, although a greater number may compromise specificity (Viswanathan et al. 1997; Gardiner and Crabb 2002). The European Glaucoma Society (EGS) therefore recommend three SAP tests per year in the two years following diagnosis (EGS 2014), although NICE guidelines are less specific, leaving this decision to clinicians' own judgement (NICE 2009; 2017). Studies have demonstrated that the true number of visual field tests carried out following diagnosis falls far short of this, at a median of 0.7 per year (Fung et al. 2013), with clinicians citing a lack of resources as a barrier to achieving this recommendation (Malik et al. 2013). Somewhat concerningly, some clinicians felt that structural assessments such as scanning laser polarimetry, or OCT, were a preferable alternative to the recommended number of perimetric tests (Malik et al. 2013); as discussed in section *1.2.5.1*, structural and functional tests are complementary, and one should not be seen as a replacement for the other. Given the variability of visual field results, particularly as disease progresses, the demonstrated sub-optimal testing intervals are likely to substantially impair progression identification.

Variability of perimetric sensitivity occurs both inter- and intra- test, and both inter- and intra- observer. Some aspects of variability may be unavoidable, e.g. the small, involuntary eye movements that occur during fixation (tremor, drift, and microsaccades) will result in a continual shift of the visual field (Martinez-Conde et al. 2004). However, it is likely that other sources of variability could be reduced, or even eliminated altogether.

The sources of variability have been studied extensively. In addition to instrument-induced variability, resulting from stimulus speed in kinetic perimetry (Goldmann 1945a; 1999), stimulus area (Gilpin et al. 1990; Wall et al. 1997; Wall et al. 2013), stimulus duration (Pennebaker et al. 1992), and background intensity (Crosswell et al. 1991), other human-related sources include anatomical variations (Armaly 1969; Wilensky and Joondeph 1984; Garway-Heath et al. 2000a), learning effects (Heijl et al. 1989b), retinal location (Heijl et al. 1987), and examiner variation in manual perimetry (Heijl and Krakau 1975). There is also some indication that verbal instructions provided by the examiner prior to conducting a visual field test may influence the results with SAP (Glen et al. 2014).

Variability also appears to occur between instrument types, sometimes because similarities in nomenclature can mask physical differences in stimuli, and care must always be taken when attempting to compare results from one instrument to another. The dB scale employed by perimeters is relative to the luminance capabilities of the instrument, '0 dB' denoting the maximum luminance available, and each dB thereafter denoting a 0.1 log unit decrease in luminance (Anderson and Patella 1999). As such, a given dB value with one instrument does not equate to the same dB value with a different instrument. This may also be true of instruments within the same perimeter series, for example the Humphrey Field Analyzer and the Octopus series', as the maximum available stimulus luminance, or even the background luminance, may differ between instruments. As such, it is recommended that patients are tested with the same instrument when monitoring for potential disease progression (NICE 2009; 2017).

A logarithmic dB scale is used as sensation relates to factors (i.e. doubling/halving intensity) rather than addition of intensity, although the use of a linear scale has been investigated, and there are indications this demonstrates a greater uniformity across perimetric sensitivity measures (Malik et al. 2006). The advantage of the dB system in SAP is that a difference of 3 dB will denote a halving, or doubling, of the luminance, irrespective of the instrument (Anderson and Patella 1999). It should be remembered, however, that logarithmic 'units' are actually unitless and merely represent a proportional change from a reference value.

1.3.3.2 Age

Much has been reported on the effect of age on threshold results, and most agree that with increasing age, there is a decrease in perimetric sensitivity (Brenton and Phelps 1986; Haas et al. 1986; Heijl et al. 1987). Many papers have subsequently attempted an age-adjustment to aid in the analysis of their results (Wilensky and Joondeph 1984; Turpin et al. 2007; Wall et al. 2013). However, as the above studies disagree with respect to the exact effect of age on perimetric sensitivity values, it is difficult to determine the accuracy of an applied age-correction.

This is especially true when considering that sensitivity reduction with age has been reported as non-uniform across the field of vision (Brenton and Phelps 1986; Haas et al. 1986; Katz and Sommer 1986; Heijl et al. 1987). Heijl et al. (1987) and Katz and Sommer (1986) reported that the reduction in sensitivity with age was eccentricity dependent, such that the hill of vision steepened, as well as flattened with age. Studies do not agree completely on the eccentricity, or quadrant, at which the greatest sensitivity difference with age is observed. Heijl et al. (1987) reported that mean sensitivity decreased more rapidly in the nasal quadrant, particularly superiorly, whereas Katz and Sommer (1986) and Haas et al. (1986) both report that the superior field demonstrates the greatest sensitivity reduction with age. Haas et al. (1986) found a greater sensitivity reduction in the centre and periphery of the visual field, and less so in the pericentral area with age, although they have not defined the boundaries of these three areas. Brenton and Phelps (1986) reported a 0.5 dB decrease per decade at fixation, a 0.6 dB decrease per decade within the central 30°, and a 0.4 dB decrease per decade between 30° and 60° from fixation.

Heijl et al. (1987), Katz and Sommer (1986), and Brenton and Phelps (1986) noted that, in addition to sensitivity reduction, the inter-observer variability of sensitivity values determined from perimetric tests increased with distance from fixation, with the greatest variability noted in the area that displayed the greatest reduction in sensitivity. Katz and Sommer (1986) noted that inter-observer variability was higher in subjects aged > 60 years, 25-30° from fixation, and a subsequent study by the same group (Katz and Sommer 1987) found that test-retest variability was also higher at locations further from fixation, and was higher in participants > 60 years. Haas et al. (1986) commented that inter-individual variability was greater in the older group, but have not indicated the limits of their age groups. Brenton and Phelps (1986) found little change in inter-individual variation from decade to decade.

It should be noted that these studies differ somewhat in their implementation. Heijl et al. (1987), Katz and Sommer (1986), and Brenton and Phelps (1986) used the Humphrey Field Analyzer, whereas Haas et al. (1986) used the Octopus Perimeter. Heijl et al. (1987) repeated perimetric tests on three separate visits, discarding results from the first session to account for learning effects. In contrast, Katz and Sommer (1986) do not report whether their subjects had experience with the Humphrey Field Analyzer, and appear to have conducted and analysed only one test; they do state that subjects had experience with other forms of perimetry, either manual, kinetic or supra-threshold, static perimetry. Brenton and Phelps (1986) tested participants who were perimetrically naïve, excluding several due to suspicious preliminary ocular examinations in an attempt to avoid including those with a genuine visual field defect. Haas et al. (1986) do not state whether subjects had experience on the Octopus Perimeter used in their study, nor whether visual fields were tested for normality prior to participants' inclusion. It is likely that these differences in study design account for some of the differences between reported findings.

One final, age-related observation in perimetric studies in general; many involve younger patients in their 20s or 30s with glaucoma (Tyler 1981; Logan and Anderson 1983; Heijl et al. 1989a; Drance 1991; Chauhan et al. 1997; Polo et al. 1998; Larrosa et al. 2000). It is likely that at least some of these individuals do not, in fact, have POAG, but perhaps pigmentary glaucoma, or even juvenile glaucoma. As different types of

glaucoma may have differing aetiologies and progression, it is possible that inclusion of differing sub-categories of glaucoma may confound results, potentially leading to inaccuracies in conclusions, particularly in studies with small sample sizes.

1.3.3.3 Software

To aid in the interpretation of visual field results, a number of software devices have been incorporated into commercial perimeters. These include: test algorithms, designed to accurately determine observer sensitivity while minimising test time, normative databases, to aid in differentiating between normal and glaucomatous sensitivity values, and analytical software, some designed to interpret one visual field test alone to identify glaucomatous test locations, and others designed to compare between tests to identify changes in sensitivity. For the purposes of this section, some of the software available for the Humprey Field Analyzer has been described, although is not intended to be an exhaustive list. Other manufacturers of commercial perimeters incorporate their own software into their instrumentation, often with the same aims as the software described here although the algorithms themselves will differ. As test algorithms employ different methods of determining perimetric sensitivity values, with differing termination criteria, this contributes to the difficulties in comparing visual field results between perimetric instruments.

Examples of software in the Humphrey Field Analyzer include the Swedish Interactive Thresholding Algorithm (SITA), Statpac, the Glaucoma Hemifield Test (GHT), and Guided Progression Analysis (GPA). SITA is an algorithm that attempts to reduce long testing times associated with Full Threshold strategies, without compromising validity of results. There are two versions, SITA Standard, and SITA Fast (Bengtsson et al. 1997a; Bengtsson and Heijl 1998a; Bengtsson and Heijl 1998b; Bengtsson et al. 1998), which are largely used as the preferred testing strategies on the Humphrey Visual Field Analyzer; it is now rare that a Full Threshold strategy would be used in a clinical setting. The SITA Standard strategy is identified as the visual field program of choice in the NICE guidelines (NICE 2009; NICE 2017).

Statpac is the normative database incorporated into the Humphrey Field Analyzer, against which sensitivity measurements are compared. This database is based on data

collected in Malmö in Sweden (Heijl et al. 1987), and Baltimore (Sommer et al. 1984), Iowa (Brenton and Phelps 1986) and Oakland (Asman and Heijl 1992) in the United States of America. The GHT was then developed from these normative results (Asman and Heijl 1992). While this software is extremely beneficial to our interpretation of visual field plots, and is regularly used in a clinical setting, there are certain limitations. For example, participants included in these studies were required to have a refractive error between ±5.00 DS (Brenton and Phelps 1986; Heijl et al. 1987; Asman and Heijl 1992), with a visual acuity of 6/9 or better, although it is a little unclear whether the subjects from Baltimore achieved 6/9 or 6/12 (Sommer et al. 1984; Asman and Heijl 1992). While these are logical limits to impose, it should be remembered that observers with refractive errors or visual acuities outside these boundaries may be classified as 'outside normal limits' in the absence of any disease. It should also be noted that the majority of participants in these studies were Caucasian, with only a minority of other ethnicities included in the Baltimore cohort (Asman and Heijl 1992). As such, the normative database and GHT may result in a higher rate of false positive or false negative outcomes when testing those of other ethnicities. Lastly, while in most cases perimetric experience was required of individuals (Sommer et al. 1984; Heijl et al. 1987; Asman and Heijl 1992), this was not the case with the data collected from Iowa (Brenton and Phelps 1986), which may result in a wider range of 'normal' sensitivities than is truly representative, due to perimetric inexperience.

Much of the software incorporated into the Humphrey Field Analyzer to aid in the analysis of perimetric data, has been designed to identify localised visual field defects; diffuse field defects will not be identified as glaucomatous (Asman and Heijl 1992). As discussed in section *1.2.5.1*, both diffuse and focal loss are apparent in most individuals with glaucoma (Drance 1991; Chauhan et al. 1997; Henson et al. 1999; Artes et al. 2005a; Artes et al. 2010). For example, the 'pattern deviation' analysis was designed to be sensitive to changes in localised field loss, but unaffected by generalised field loss, by correcting the field height ('total deviation' analysis) for generalised depression of sensitivity; the 85th percentile (7th most positive value, excluding the three test locations adjacent to the blind spot) is denoted as '0', and the remaining test locations are assigned a value relative to this (Bengtsson et al. 1997b).

In doing so, the intention is to distinguish between optical deviation, e.g. media opacities or optical defocus, expected to cause a general decrease in sensitivity across the entire visual field, and those caused by glaucoma, expected to cause a focal-type loss. However, as almost all glaucomatous visual field progression has been found to comprise both focal and diffuse components (Henson et al. 1999; Artes et al. 2005a; Artes et al. 2010), and the total deviation analysis reportedly identifies glaucoma earlier than the pattern deviation analysis (Artes et al. 2005a), software designed to identify focal loss only may mask glaucomatous changes in the visual field.

GPA uses the first two tests undertaken by the observer as a 'baseline' sensitivity measure. In subsequent tests, perimetric sensitivity at each location is compared with its baseline value to determine whether the current perimetric sensitivity value represents a significant change. Differences occurring in two consecutive tests, and in three or more test locations, are flagged as indicative of 'possible progression', or 'likely progression' if identified in three consecutive tests. Artes et al. (2014) noted that, as GPA is based on an average test-retest variability, it is likely to underestimate progression in those individuals with low variability, and overestimate in those with high variability. They tested 30 patients with glaucoma twelve times over twelve weeks, and observed a false positive rate (i.e. locations flagged by GPA as progressing) of 18.5 % for 'possible progression', and 2.6% for 'likely progression'. They noted a large variation between individuals, observing a correlation between false positives on GPA and higher-than-average reliability indices, which may aid in the identification of those individuals with a false positive GPA result in a clinical setting. Nouri-Mahdavi et al. (2011) noted that, due to the high test-retest variability observed in moderate to advanced disease, GPA will not comment on locations with lower sensitivity, instead indicating these areas with an 'X'. As such, while GPA may aid in the identification of visual field progression, it cannot be relied upon solely, but in conjunction with clinical expertise.

1.3.4   Selective versus non-selective perimetry

In an attempt to increase the reliability of perimetry findings, studies have investigated whether glaucoma preferentially damages some types of retinal ganglion cells earlier than others. If this occurs as part of the pathogenesis of glaucoma, psychophysical

tests designed to selectively evaluate those specific ganglion cells could be developed, resulting in earlier detection, and more effective monitoring, of glaucoma (Johnson 1994).

It is widely accepted that a 'parallel processing' takes place in the visual system, in which different subsets of retinal ganglion cells project to different layers of the LGN within separate processing streams. These are generally known as the 'parvocellular', 'magnocellular', and 'koniocellular' pathways (Casagrande 1994; Lee 1996), and each are considered as specialised towards different functions. The parvocellular pathway is constructed of smaller diameter fibres, and is preferentially stimulated by colour, pattern, and acuity, while the magnocellular pathway is made up of larger diameter fibres, and thought to be involved in the assessment of motion and spatial relationships between objects (Merigan and Maunsell 1993; Johnson 1994). The koniocellular pathway is preferentially stimulated by colour at the short-wavelength (blue) end of the spectrum (Martin et al. 1997).

Various studies have reported that glaucoma causes a selective loss of the retinal ganglion cells from one pathway or another. Perhaps most notoriously, Quigley et al. (1987; 1988) reported that the large-diameter ganglion cell axons were selectively damaged in glaucoma. This, coupled with reports that glaucomatous damage could be identified earlier with the use of flicker stimuli (Tyler 1981), led to perimetric tests designed to selectively stimulate the ganglion cells in the magnocellular pathway. These included Temporal Modulation (flicker) Perimetry, Motion Displacement Perimetry, and Frequency Doubling Technology (FDT).

Contradictory reports of impaired colour vision in patients with glaucoma and OHT (Kalmus et al. 1974; Poinoosawmy et al. 1980; Adams et al. 1982) prompted the exploration of colour perimetry (Logan and Anderson 1983), and specifically short-wavelength sensitivity. This resulted in the well-known Short-Wavelength Automated Perimetry (SWAP), designed to selectively stimulate the retinal ganglion cells in the koniocellular pathway.

1.3.4.1 Non-selectivity

Given that initial reports on 'selective' types of perimetry, such as SWAP (Johnson et al. 1993a; 1993b), or FDT (Maddess and Henry 1992) indicated earlier detection of glaucomatous damage compared to SAP, irrespective of the pathway targeted, the 'reduced redundancy' hypothesis was posed (Johnson 1994) as an alternative to the 'selective loss' hypothesis. Johnson (1994) proposed that, by targeting a sparse population of retinal ganglion cells, i.e. with minimal overlap and redundancy, it may be possible to reveal earlier functional losses, even if the subpopulation of cells has not been preferentially damaged. This would explain why perimetric tests that aimed to selectively stimulate different subpopulations of retinal ganglion cell, all appeared to show a greater sensitivity to glaucoma than SAP.

Johnson (1994) also raised some concerns with respect to the conclusions drawn by Quigley et al. (1987; 1988) regarding selective loss. He noted that, while results appeared to show a reduction of large diameter nerve fibres, this was not apparent in mild glaucomatous damage; this being the case, the conclusion that these fibres are selectively targeted could be somewhat misleading. Johnson (1994) acknowledged that the finding was important, but not necessarily to the exclusion of other factors, or to the sole criterion in the development of new clinical test procedures. Additionally, inconsistencies in the remaining cell diameters in glaucomatous eyes are apparent in the study of Quigley et al. (1989), in which some retinal areas had a significantly smaller cell diameter than normal, some had a significantly larger cell diameter than normal, and some were not significantly different, with no apparent correlation to disease severity.

A more recent histological study examining retinal ganglion cells has indicated a general ganglion cell *shrinkage*, rather than a selective loss of larger diameter cells, in glaucoma (Morgan et al. 2000). In addition, histological studies examining the LGN and V1 in the visual cortex of monkeys with experimental glaucoma, suggest that, if anything, the P-cells of the parvocellular pathway are more affected by glaucoma than the M-cells of the magnocellular pathway (Crawford et al. 2000; Yücel et al. 2000).

Furthermore, a thorough review of magnocellular and parvocellular pathways indicates that, while the two pathways are anatomically distinct, this is not so for their physiological responses (Merigan and Maunsell 1993). For example, although the parvocellular pathway is considered the pathway for colour processing, the magnocellular pathway does show a non-selective response to colour transitions. Of particular relevance to perimetric testing is that cells of both the magnocellular and parvocellular pathways have different, but *largely overlapping,* responses to spatial, temporal and luminance characteristics.

This is supported by the findings of Ennis and Anderson (2000), in their investigation of thresholds using flickering gratings of differing temporal frequency, under different levels of background luminance. They did not find a discrete 'break' in the curves, but rather a gradual change in the number of cells responding, suggesting the absence of a sudden shift from the parvocellular to the magnocellular pathway, as would be expected if the physiological responses of each pathway were distinctly separate. Thus, in perimetric tests designed to preferentially stimulate one particular pathway, one must consider whether the desired pathway has been truly isolated.

Two of the most popular perimetric tests resulting from explorations of selectivity are discussed in more detail here, namely SWAP and FDT.

Short-Wavelength Automated Perimetry (SWAP)

Since the 1980s and 1990s, much attention has been given to short-wavelength sensitivity as a means of detecting early glaucomatous change, resulting in the incorporation of SWAP technology into commercially available perimeters (Haag-Streit 2014; Zeiss 2014; Zeiss 2015). In SWAP, a bright yellow background depresses the sensitivity of the medium- and long- wavelength cones, and a blue target preferentially stimulates the short-wavelength cones (Heron et al. 1988; Wild 2001). This method of preferentially stimulating the koniocellular pathway is an established psychophysical technique. Many studies have investigated this strategy, suggesting that SWAP can detect glaucomatous damage earlier than white-on-white perimetry (Heron et al. 1988; Johnson et al. 1993a; Johnson et al. 1993b; Johnson et al. 1995; Wild 2001). However, substantial limitations have been identified.

For example, before the inclusion of SWAP into commercially available perimeters, different parameters of luminance, and stimulus duration were often used in differing studies (Heron et al. 1988; Johnson et al. 1993a; Johnson et al. 1993b; Larrosa et al. 2000). This, coupled with the lack of a standardised normative database (Wild 2001), created difficulties in drawing accurate conclusions and comparing results between studies.

Previous perimetric experience and learning effects were often not considered (Johnson et al. 1993a; 1993b). This has particular implications given that Wild and Moss (1996) found that the learning effect for SWAP was independent of previous perimetric experience with white-on-white perimetry, i.e. SWAP sensitivities were no less variable in individuals with prior experience in white-on-white perimetry, than in those with no perimetric experience.

Both ocular media and macular pigment absorb short-wavelength light, which must be corrected for when interpreting findings from SWAP; this varies greatly between individuals and so must be established for each observer, which can be a difficult and time-consuming endeavour in a clinical setting (Johnson et al. 1993a; Johnson et al. 1993b; Johnson et al. 1995; Johnson 1996; Wild et al. 1998; Larrosa et al. 2000; Caprioli 2001; Wild 2001). Given that glaucoma investigations are typically undertaken in a population demographic affected by lenticular yellowing and/or opacities, ocular media absorption is likely to be substantial.

Despite extensive studies on test-retest variability with white-on-white perimetry, studies often did not account for this with SWAP (Johnson et al. 1993a; Johnson et al. 1993b; Johnson et al. 1995). Participants often undertook only one SWAP test, and visual field abnormality was defined as threshold deviation from normal, without accounting for variability of the threshold estimate (Wild 2001). Indeed, test-retest variability is apparent in the example visual field plots of Johnson et al. (1993a), although no comment has been made. This is of particular concern as studies have since indicated that both short-term and long- term fluctuations, in addition to inter-individual variability, are significantly higher with SWAP than with white-on-white perimetry (Wild and Moss 1996; Kwon et al. 1998). As such, a substantially greater

reduction in sensitivity must be clinically demonstrated before a diagnosis of glaucoma can be confirmed.

The dynamic range is more limited with SWAP than white-on-white perimetry, as a higher background luminance is necessary to depress the sensitivity of the medium- and long- wavelength cones. The maximum stimulus increment luminance available in SWAP is 20.6 cd/m^2, in comparison with 3183.1 cd/m^2 in white-on-white perimetry; SWAP is therefore not suitable in cases of moderate-advanced visual field loss (Johnson et al. 1993b; Wild 2001).

It should also be noted that, despite development from an established psychophysical technique, known to isolate the short-wavelength sensitive pathway in those with normal vision or mild disease, there are indications that responses may be elicited by alternative pathways in those with advanced disease (Felius et al. 1995; Demirel and Johnson 2000; Wild 2001).

To summarise, although developed from an established psychophysical method, SWAP has not translated effectively to a clinical setting due to the substantial limitations involved, and has largely fallen out of clinical use.

Frequency doubling Technology (FDT)

'Frequency doubling' is the term given to a visual phenomenon in which a stimulus of low spatial frequency undergoes high temporal frequency counterphase flicker over a certain frequency range, thus giving the illusion that the spatial frequency of the stimulus has doubled (Kelly 1966). These frequency characteristics were believed to be identified solely by a non-linearly responding subset of the magnocellular ganglion cells (M_y cells).

As previously noted, Quigley et al. (1987; 1988) had reported that large diameter nerve fibres, thought to be specific to the magnocellular pathway, were selectively damaged in glaucoma, and Dandona et al. (1991) found a decrease in magnocellular layers of the dorsal LGN in the presence of glaucoma. This led to the development of perimetric techniques attempting to utilise the frequency doubling phenomenon, with the hypothesis that the use of such a phenomenon would identify visual field loss earlier

than SAP by preferentially stimulating the ganglion cells of the magnocellular pathway, and particularly the subgroup M_y cells (Maddess and Henry 1992; Johnson and Samuels 1997). More recently however, the presence of this subgroup has been called into question (White et al. 2002).

FDT perimetry largely employs a target of 10° x 10°, with a 5° x 5° central stimulus. The stimulus is a sinusoidal grating of 0.25 cycles per degree, undergoing 25-Hz counterphase flicker (Cello et al. 2000). The successor to FDT, FDT2 (Matrix), differs slightly by employing sinusoidal grating stimuli of 0.50 cycles per degree, each with a square window of 5° x 5°, undergoing 18-Hz counterphase flicker (Anderson et al. 2005).

Although initial tests attempted to ensure the frequency doubling illusion had been identified by observers (Maddess and Henry 1992), this was not the case in later testing of the technique, which would become the FDT strategy (Johnson and Samuels 1997). The task incorporated in the FDT perimeter is intended to establish threshold contrast of the grating stimulus once the frequency doubling illusion is generated (Allen et al. 2002), however, as no confirmation of this illusion is required by the observer, there is no guarantee of its perception. The term 'frequency doubling perimetry' is therefore somewhat misleading.

There is some disagreement as to whether FDT detects visual field damage earlier than SAP, although most studies seem to agree that the results between the two are fairly comparable (Johnson and Samuels 1997; Chauhan and Johnson 1999; Artes et al. 2005b; Haymes et al. 2005; Liu et al. 2011; Redmond et al. 2013a). Valid comparisons between these two techniques must appropriately overcome the difficulties associated with analysing sensitivity values obtained on different testing platforms with differing dB scales, particularly as these two techniques employ very different stimulus types. It is likely that this accounts for some of the disagreement between studies. A number of different analytical methods have been attempted to achieve this objective.

A longitudinal study by Haymes et al. (2005), reported differing results with differing analytical techniques. Analysis with the 'glaucoma change probability' technique, as employed in the Humphrey Field Analyzer, indicated that FDT identified visual field

defects earlier than SAP, however with a linear regression analysis, the opposite was found, and SAP identified defects earlier than FDT. As glaucoma change probability was developed for use with the Goldmann III stimulus in white-on-white perimetry, it may be that its use in FDT analysis is not appropriate, due to the differing dB scales. Limitations in this study included the lack of normal control participants, such that the rate of false positive 'progression' was not determined, and participants' lack of experience with FDT perimetry; it cannot be assumed that experience with SAP is transferrable to FDT.

Artes and Chauhan (2009) converted threshold results into a signal/noise ratio (SNR) in their comparison of FDT2 with SAP. As the instrument-specific units were in both the numerator and the denominator, they cancelled each other out, such that the resulting SNR was independent of the dB scale. They noted that both signal and noise appeared numerically larger with FDT2 compared with SAP, resulting in a similarity between the SNR of both techniques. They did observe a higher SNR with FDT2 when comparing between superior and inferior sectors, using a similar method as utilised in the GHT, thus indicating an overall gain with FDT2.

Redmond et al. (2013a) used a newer analytical technique known as 'permutation of pointwise linear regression' (PoPLR) to investigate longitudinal data. This method is well suited for comparison between techniques that differ in stimulus size and scale, as it is individualised to each observer rather than relying on population-based normative values. They found that FDT2 identified deterioration in fewer glaucoma patients than SAP when comparing total deviations, and approximately equal deterioration when comparing pattern deviations. As such, they concluded that FDT2 did not appear to be more sensitive in identifying visual field deterioration than SAP.

Studies agree that test time with FDT is substantially reduced in comparison with SAP, which is certainly an advantage (Johnson and Samuels 1997; Chauhan and Johnson 1999). Test-retest variability is also reported as generally more uniform across the range of perimetric sensitivity values than with SAP (Chauhan and Johnson 1999; Artes et al. 2005b; Fredette et al. 2015). These studies showed similar test-retest values for SAP and FDT at high sensitivity values, however FDT test-retest variability appeared

more uniform with sensitivity, and did not appear to display substantially greater test-retest values at lower sensitivities. This finding could be explained by the differing dB scales between the two techniques.

Of these studies, that of Chauhan and Johnson (1999) was conducted using original FDT parameters, and it is not stated whether participants had adequate perimetric experience. Unusually, patients with pseudoexfoliative glaucoma were included in this study; as this sub-type of glaucoma has been shown to progress faster than POAG (Heijl et al. 2009), it is often excluded in studies of this nature, as results may be unduly influenced. The study by Artes et al. (2005a) was conducted using FDT2, and subjects had perimetric experience with both SAP and FDT. In this study, only participants with glaucoma were tested, with no normal controls included. As such, it is perhaps unexpected that similar conclusions were drawn in both studies.

FDT has been found to be more robust to optical defocus at the fovea, although smaller stimuli, such as those used in FDT2, were less robust (Anderson and Johnson 2003). Artes et al. (2003b), examining a greater number of test locations out to 30° with the FDT perimeter, found a small reduction in perimetric sensitivity due to blur. Contrary to this, FDT has been found to be less robust to the effects of reduced retinal illuminance, as caused by pupil miosis or an increase in intraocular straylight (Johnson and Samuels 1997; Swanson et al. 2005; Anderson et al. 2009; Bergin et al. 2011). This may be a substantial disadvantage given the population demographic in which perimetry is used most, in which pupil miosis and/or cataracts are frequently observed.

There is also evidence to suggest that the Goldmann III stimulus, as used in SAP, is actually superior to the grating stimulus incorporated in FDT in preferentially stimulating the magnocellular pathway (Swanson et al. 2011).

Given that FDT incorporates a very different stimulus to that of SAP, it is difficult to compare results between the two techniques in clinical practice; those with established glaucoma may have been monitored for several years with the well-established SAP technique, and it is generally recommended that patients are monitored consistently with the same testing strategy (NICE 2009; 2017). FDT is

subsequently less understood among clinicians, which could also create problems if attempting to change patient monitoring from one type of perimetry to another. Given the findings detailed above, however, there may be little advantage in using FDT in clinical practice.

More recently, studies have examined the potential of exploiting spatial summation changes in perimetry as an alternative to SAP. This is discussed in section *1.4*.

1.4   Spatial Summation

Recent studies examining spatial summation in glaucoma have indicated that a perimetric test designed with the purpose of identifying spatial summation changes may hold promise as an alternative to SAP. 'Spatial summation' describes the method by which the visual system integrates the amount of light energy over a given stimulus area (Redmond and Anderson 2011). This 'summing' of information is fundamental to the way in which the visual system processes information.

Ricco's law states that, for a range of small stimuli, threshold intensity is inversely proportional to the area of the stimulus at threshold, as per *Equation 1.2*, giving a slope of -1 when stimulus intensity at threshold is plotted against stimulus area. The stimulus is said to undergo 'complete spatial summation', and Ricco's area is the largest stimulus area for which this law holds (Ricco 1877). Beyond this area, summation is only partial, also referred to as 'incomplete' or 'probability' summation (*Figure 1.7*). A Goldmann III stimulus, as is currently utilised in SAP, is larger than Ricco's area within the central 15° of the visual field, thus undergoing incomplete spatial summation (Swanson et al. 2004).

$$I \times A = k$$

$I = Intensity\ (\Delta I/I)$

$A = Area\ (Deg^2)$

$k = constant$
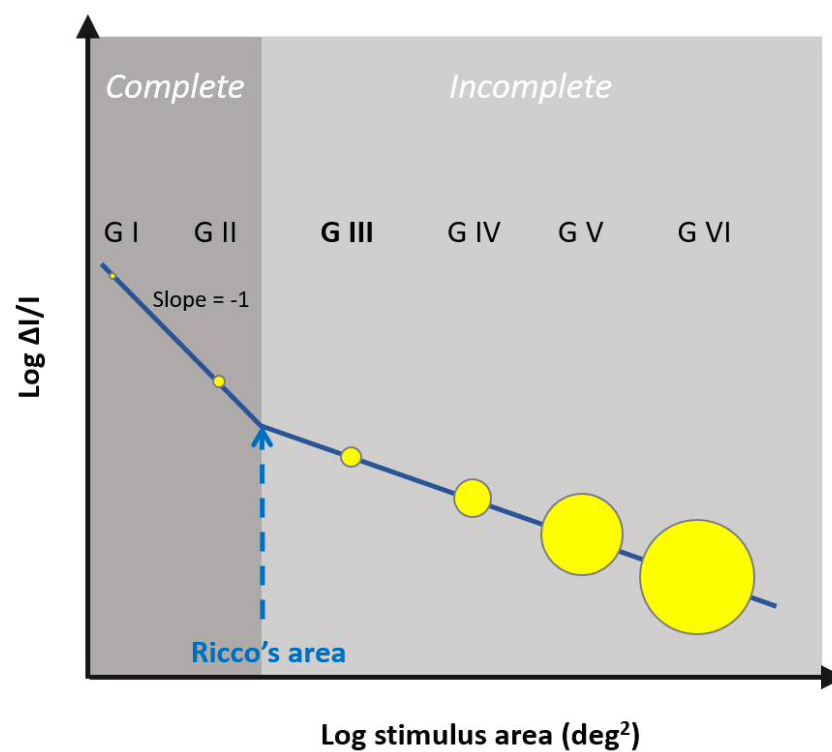
Figure 1.7 – The threshold intensity of Goldmann I-VI, forming the spatial summation curve. Ricco's area, the limit of complete spatial summation, is shown. 'G III' Indicates the area of a standard Goldmann III stimulus relative to the spatial summation curve within the central 15° of the visual field.

1.4.1   Complete spatial summation

Ricco's area is not a fixed value; there is high inter-individual variance (Schefrin et al. 1998; Redmond et al. 2010b), and many factors have been shown to affect its size. Several studies have noted that Ricco's area is greater with increasing eccentricity from fixation (Graham and Bartlett 1939; Graham et al. 1939; Hallett 1963; Wilson 1970; Vassilev et al. 2003). Latham et al. (1993) noted that the change in perimetric

sensitivity with eccentricity occurred more slowly than was suggested by the rapid decline in ganglion cell density. They suggested that this was due to the increase in ganglion cell receptive field area, and thus the change in spatial summation with eccentricity from fixation. A lower Ricco's area has been reported with an increase in background intensity, and with an increase in stimulus duration (Barlow 1957; Barlow 1958; Davila and Geisler 1991). Ricco's area measurements have also been observed to differ with wavelength of the stimulus, increasing with shorter wavelengths, and particularly in isolation of the S-cone pathway (Hallett 1963; Felius et al. 1997; Vassilev et al. 2000). An enlarged Ricco's area has also been reported with an increase in accommodation and convergence (Richards 1967).

'Temporal summation' describes the method by which the visual system integrates the amount of light energy over time, and bears many similarities to that of spatial summation, referred to as 'complete temporal summation' when stimulus intensity and stimulus duration are inversely proportional at threshold, i.e. Bloch's law (Bloch 1885), and 'incomplete temporal summation' when this relationship no longer holds. The limit of complete temporal summation has been found to vary under similar conditions as that of the limit of complete spatial summation, e.g. with background intensity (Barlow 1957; 1958). Indeed, Piéron (1920) noted a reciprocal relationship between stimulus duration and complete spatial summation, and stimulus area and complete temporal summation.

1.4.2   Incomplete spatial summation

A number of laws have been proposed to describe the relationship between stimulus area and stimulus intensity beyond that of Ricco's area. Piper's law states that intensity of a stimulus at threshold is inversely proportional to the square root of the stimulus area, as per *Equation 1.3* (Piper 1903). This corresponds to a slope of -0.5 when log threshold intensity is plotted against log stimulus area.

$$I \times \sqrt{A} = k$$

$I = Intensity\ (\Delta I/I)$

$A = Area\ (Deg^2)$

$k = constant$

Pieron's law states that intensity of a stimulus at threshold is inversely proportional to the cube root of the stimulus area, as per *Equation 1.4* (Piéron 1929). This corresponds to a slope of -0.3 when threshold intensity is plotted against stimulus area.

$$I \times \sqrt[3]{A} = k$$

$I = Intensity\ (\Delta I/I)$

$A = Area\ (Deg^2)$

$k = constant$

Goldmann (1945b; 1999), noted that a fourfold increase in area was approximately equivalent to an increase in intensity by a factor of 3.16, and calculated a summation coefficient of approximately 0.8, sometimes referred to as 'Goldmann's approximation'. He found that this coefficient was lower with larger stimuli, and also that it differed in different regions of the visual field, however concluded that for the purposes of perimetry, a constant value of 0.8 should be assumed, and used this to determine the area and intensity values in the Goldmann perimeter.

With sufficiently large stimuli, intensity becomes independent of area (Löhle 1929; Graham et al. 1939), such that the slope of the summation function approaches zero (Weber's law). Wilson (1970) noted that no law, either for complete or partial summation, holds for a substantial range of stimulus areas, and suggested that, as no single slope adequately described incomplete summation, this was best fitted with a curve of steadily decreasing slope, a schematic of which is shown in *Figure 1.8*.
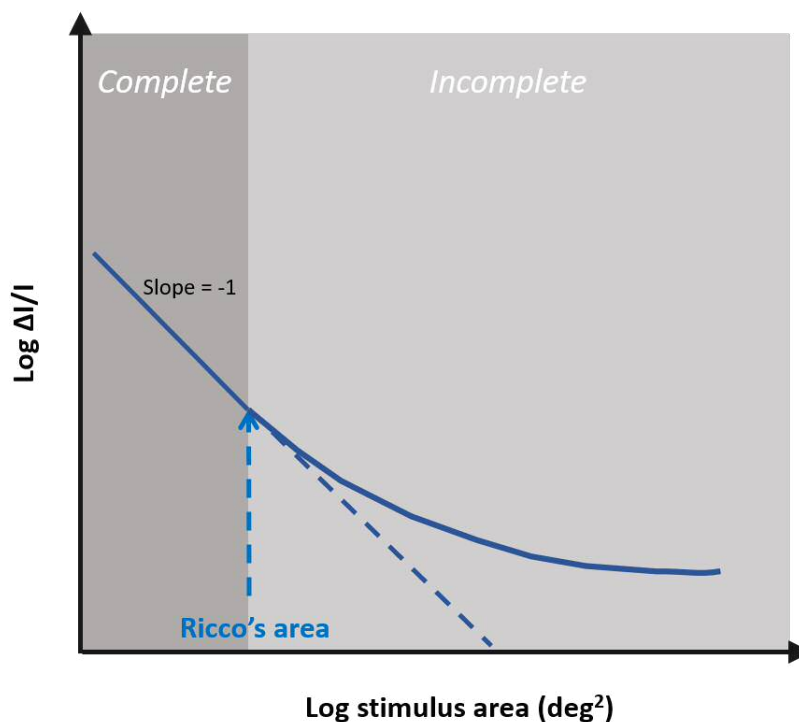
Figure 1.8 – Spatial summation curve as proposed by Wilson (1970), in which partial summation is fitted with a curve of steadily decreasing slope.

1.4.3   Underlying physiology governing spatial summation

The relationship between stimulus area and intensity for just-visible stimuli has formed the basis of many attempts to explain the functional organisation of the visual system (Hallett 1963). The dispute about what Ricco's area, the limit of complete spatial summation, is actually describing, and where in the visual system it is governed, is still ongoing. One of the more popular hypotheses stated that the size of Ricco's area

describes the structure of the retina, specifically the receptive fields of the retinal ganglion cells.

1.4.3.1 Spatial summation at the retinal level

It is well established that a 'convergence' takes place in the retina of photoreceptors to retinal ganglion cells. In recording responses from single ganglion cell axons, Hartline (1938) noted that a response was only elicited by illuminating a specific retinal location, termed the receptive field of the ganglion cell. In a later study, Hartline (1940) also noted that sensitivity to light was not uniform across the whole receptive field, and that a lower threshold, and a greater response, was ellicited at the receptive field centre. Other studies have noted an annular surround of the receptive field that inhibits the central response (Barlow 1953; Kuffler 1953), which may lead to the occurrence of a larger stimulus evoking a weaker response than a smaller one. It was therefore hypothesised that Ricco's area represented the area of the receptive field centre, and that Ricco's law no longer held for larger stimuli due to the antagonistic nature of the annular surround (Glezer 1965; Ikeda and Wright 1972).

Further support for this hypothesis came from studies that observed differences in Ricco's area that echoed differences in receptive field centre. As noted in section *1.4.1*, several studies have reported greater values for Ricco's area with increasing retinal eccentricity from fixation, and it is well-established that receptive field centre size also increases with eccentricity from the fovea, in accordance with an increased convergence from photoreceptors to ganglion cells; at the fovea there may be a ratio of one photoreceptor to one ganglion cell, and in the peripheral retina this ratio may increase to several hundred photoreceptors to one ganglion cell (Hartline 1940; Wiesel 1960; Jacobs 1969). Hartline (1938) noted that, while the *location* of the receptive field was fixed, the *extent* was not, and was dependent upon the intensity and size of the stimulus used, and on the state of light adaptation.

Hallett (1963) hypothesised that there were at least two types of spatial organisation of rod photoreceptors, and that the spatial summation curve may represent not just one, but several different summation areas, perhaps explaining why several different laws exist to describe partial summation, as described in section *1.4.2*. Volbrecht et al.

(2000a) showed that retinal ganglion cell density, rather than cone density, more closely defined Ricco's area in the parafoveal and peripheral retina, but had difficulty concluding which type of ganglion cell was contributing to their findings. A series of studies by Vassilev et al. (2000; 2003; 2005) on the S-cone pathway provided evidence that Ricco's area was dependent on the dendritic field diameter of retinal ganglion cells, particularly the small bistratified ganglion cells under S-cone isolation, although within this series of studies, some findings indicated a closer correlation between Ricco's area and dendritic field diameter than others. Both Volbrecht et al. (2000b) and Vassilev et al. (2003) reported that the increase in Ricco's area measurements with eccentricity occurred at a faster rate than that of the dendritic field diameter for small bistratified retinal ganglion cells, and Felius et al. (1997) noted that Ricco's area measurements for their normal, control subjects were larger than the dendritic field size of retinal ganglion cells, as reported by anatomical studies.

With evidence of a close, but not exact, correlation between Ricco's area and retinal ganglion cell receptive field size, various studies have investigated other potential influences of spatial summation along the visual pathway.

1.4.3.2 Spatial summation along the full visual pathway

Davila and Geisler (1991) noted that there are three main levels at which spatial summation may occur along the visual pathway. The first level relates to the optics of the eye, where pupil diffraction and optical aberrations degrade the stimulus image, the second level relates to the summing of information from photoreceptors, and the third level relates to neural summation. The popular theory that Ricco's area describes the receptive field centre of the retinal ganglion cells, as described in section *1.4.3.1*, relates to summation at the level of the photoreceptors, but summation at the other two levels, or the influence these levels may have on Ricco's area measurements, must also be considered.

Davila and Geisler (1991) measured Ricco's area at the fovea, and compared these measurements with that of an ideal observer, in which photoreceptor summation was not a factor. Their measurements were very similar to that of an ideal observer, such that they advocated that Ricco's area at the fovea is largely/completely accounted for

by preneural factors, particularly in photopic conditions in which there was closer agreement with that of an ideal observer. They noted that their findings were in agreement with anatomical studies, in which a 1:1 ratio of photoreceptors to ganglion cells was observed at the fovea (Boycott et al. 1969; Jacobs 1969), although conceded that summation of up to two cones may be undetectable. Under scotopic conditions, as noted in other studies, Ricco's area measurements were greater, suggesting summation of photoreceptors (Glezer 1965; Lie 1981). While this finding indicates that preneural factors substantially contribute to Ricco's area, given that only the fovea was tested, it is difficult to ascertain whether this finding supports the hypothesis that Ricco's area describes the central receptive field of a ganglion cell (an area of perhaps a single cone at the fovea), or contradicts it. Peripheral measurements, in which there is a much greater known ratio of photoreceptors to ganglion cells, may be beneficial in this endeavour.

Dalimier and Dainty (2010) also investigated preneural factors by using an adaptive optics vision simulator to correct ocular aberrations, before determining Ricco's area at the fovea, and comparing these measurements with those obtained without correction of optical aberrations. Their data were largely in agreement with that of Davila and Geisler (1991) in demonstrating the substantial role played by optical aberrations in both threshold and Ricco's area measurements, which they attributed to the increased optical spread at the level of the retina (the 'pseudosummation' area). However, they concluded that while optics influenced Ricco's area measurements, they did not fully account for the measurements at the fovea, particularly given the findings of Davila and Geisler (1991) under lower light levels. Other studies, although not primarily measuring optical factors, have independently concluded that their results could not have been explained by preneural factors alone, and were not sufficiently altered by applied optical corrections (Schefrin et al. 1998; Volbrecht et al. 2000a).

Schefrin et al. (1998) found a greater Ricco's area with increasing age under scotopic conditions. Their analysis determined that this could not be explained by degradations due to optical factors alone, but indicated some form of neural change. Their comparison of Ricco's area measurements, coupled with rod and ganglion cell counts

from two other studies (Curcio et al. 1993; Curcio and Drucker 1993), indicated that some kind of synaptic rewiring, either at the level of the retina or at the level of the visual cortex, resulting in a greater convergence of photoreceptors to ganglion cells/ganglion cells to cortical space, could explain their findings of a greater Ricco's area with age. This concept of synaptic rewiring, at either the level of the retina or the visual cortex, had previously been put forth by Richards (1967) more than thirty years prior, as an explanation for observed changes in receptive field size.

Latham et al. (1994) found that spatial summation curves for younger and older observers could be superimposed neatly by displacing the curves along the sensitivity axis only, and Redmond et al. (2010b) found no difference in Ricco's area with increasing age under photopic conditions. This may in part be due to the stability of cone numbers with age, in comparison to those of rods (Curcio et al. 1993), however significantly lower retinal ganglion cell numbers have been demonstrated across the central 11° of the retina with an increase in age (Curcio and Drucker 1993). As such, a difference in Ricco's area with increasing age would also be expected under photopic conditions if Ricco's area were determined by retinal factors alone. A follow-up study by Redmond et al. (2011) found a lower Ricco's area with increased intraocular straylight, noting that it appeared to be equal and opposite to that expected for a decline in ganglion cell density over a given age range. They hypothesised that this could be the reason they had failed to identify a difference in Ricco's area with age in their previous study.

Schefrin et al. (2004), examining the onset of Weber-like behaviour (i.e. a constant log contrast sensitivity with increasing retinal illuminance) using two gratings of differing spatial frequency, did not find a significant difference with age with either grating, concluding that the receptive field centres of retinal ganglion cells receiving input from rod photoreceptors did not enlarge with age. Thus, the synaptic re-wiring they had proposed to explain a greater Ricco's area with increasing age in their previous study (Schefrin et al. 1998) could not be retinal. They did find a reduction in contrast sensitivity with age for the grating of lower spatial frequency, but not for the grating of higher spatial frequency. This was consistent with some previous studies (Schefrin et al. 1999, Hennelly et al. 1998), although another study noted the opposite effect

(Owsley et al. 1983), with a decrease in contrast sensitivity with age noted for higher spatial frequencies, while remaining largely unaffected for lower spatial frequencies. All studies concluded that these findings were unlikely to be explained by optical factors alone, and likely also represented neural change. Schefrin et al. (2004) therefore concluded that their previously observed difference in Ricco's area measurements across different age groups under scotopic conditions must be influenced at the level of the visual cortex.

Although the concept of cortical involvement is not new, there is increasing evidence to suggest its involvement in the determination of Ricco's area. Je et al. (2018) have noted a larger Ricco's area in amblyopic versus fellow non-amblyopic eyes, and given that amblyopia is accepted as a largely cortical phenomenon (Sengpiel and Blakemore 1996), this finding lends support to the concept of a cortical influence on Ricco's area measurements. Gilbert and Wiesel (1992) demonstrated the plasticity of the adult visual cortex in response to retinal changes by recording responses of the mammalian primary visual cortex before and after the application of retinal lesions. A substantial increase in receptive field area was noted in those receptive fields located close to the retinal lesion boundary in the immediate aftermath of the lesion formation. In addition, although receptive fields located near the centre of the lesion were unresponsive in the immediate aftermath, a re-mapping of these cortical areas was apparent in repeated readings over a two-month period, at the end of which activation of all cortical areas could be achieved, with observed shifts in field position of up to 5°. This plasticity did not appear to occur all along the visual pathway, as areas of the LGN were still unresponsive at the end of the two-month period.

The adaptations in the visual cortex to changes at the retinal level lend support to the hypothesis that Ricco's area could be influenced at the cortical level, given the immediacy of the demonstrated adaptations. Since an enlarged Ricco's area has also been found in the presence of retinal disease (Sloan 1961; Sloan and Brown 1962; Dannheim and Drance 1974; Redmond et al. 2010a), which is discussed in more detail in section *1.4.4*, it is possible the cortical reorganisation described by Gilbert and Wiesel (1992), both short-term and longer- term, explains the mechanism by which this enlargement occurs.

While more traditional theories advocated that enlargements in Ricco's area occurred to maintain stimulation of an equal number of retinal ganglion cells, for example across different retinal eccentricities (Fischer 1973), the indications of cortical involvement have resulted in expansions on this theory, with suggestions that differences in Ricco's area may occur to maintain a constant cortical space (Ransom-Hogg and Spillmann 1980; Redmond et al. 2010a). Swanson et al. (2004) developed a two-phase 'hockey-stick' model based on complete and incomplete spatial summation, relating log ganglion cell numbers to visual field sensitivity in dB. This model provided a close fit, accounting for 82% of their data, and indicated that there were 31 retinal ganglion cells underlying Ricco's area at each retinal location.

As demonstrated here, there is substantial evidence in the literature to suggest the influence of all three levels along the visual pathway, as proposed by Davila and Geisler (1991), in the determination of Ricco's area. Although it is likely that the traditionally advocated retinal receptive field size does play a role, it should be remembered that the structural, dendritic field, and the functional, perceptive field, are not the same (Anderson 2006). As such, and as proposed by Wilson (1970), it is likely that no one area of the visual pathway completely governs the size of Ricco's area, but that Ricco's area reflects an accumulation of factors at all levels of the visual pathway.

1.4.4 Spatial summation as an indicator of disease

Although the use of spatial summation to identify disease, particularly glaucoma, has enjoyed something of a renaissance in recent years (Redmond et al. 2010a; Kalloniatis and Khuu 2016; Phu et al. 2017b), it is not a new concept. Sloan (1961) investigated the relationship between area and threshold luminance for stimuli in the centre and periphery of the retina, and noted that the relationship between the two was different in those with certain visual defects compared with normal controls. Threshold values for different stimulus areas were plotted for subjects with a range of retinal diseases, and appeared to indicate that disease affecting the optic nerve resulted in an increase in spatial summation.

Dannheim and Drance (1974) had previously attempted to investigate spatial summation differences between normal observers and those with glaucoma. They

noted a difference in spatial summation in glaucomatous visual field locations in comparison to a normal area of the same eye at an equidistant location from fixation, however reported that this was apparent in less than half of the eyes tested. There are several possibilities which may explain why a difference in spatial summation was not apparent in all subjects, given the findings of other studies (Redmond et al. 2010a; Kalloniatis and Khuu 2016; Phu et al. 2017b). The authors noted some difficulties in obtaining measurements, commenting that narrow field defects were difficult to map, and there was substantial variability of some threshold measurements, potentially confounding the establishment of Ricco's area. The glaucomatous field locations were compared to apparently normal locations in the same eye, however it is possible that the areas identified were not truly 'normal', and may have represented a subtler glaucomatous defect, particularly as these locations were identified with kinetic perimetry, in which sensitivities have been found to be highly dependent on the perimetrist (Trobe et al. 1980). As such, differences in spatial summation were likely much smaller, and too subtle to identify. Additionally, the study was carried out at two background luminances, 3.2 and 0.03 cd/m^2, which are lower than is commonly used in studies of this type (10 cd/m^2), and at a substantially greater stimulus duration, one second, than is generally used (0.1 to 0.2 seconds). As Ricco's area is known to be smaller with both a lower background luminance, and a longer stimulus duration (Barlow 1957; Barlow 1958; Davila and Geisler 1991), it is possible that there is less of a difference between normal and glaucomatous spatial summation curves under these conditions, which may explain why this study failed to identify differences in spatial summation in all eyes tested. Indeed, Fellman et al. (1988) reported that differences in threshold luminance due to changes in background luminance differed between those with glaucoma and normal, control subjects, and that this difference was dependent on eccentricity from fixation. As Dannheim and Drance (1974) observed locations at different eccentricities for different subjects, this may explain why differences in spatial summation from normal were noted at some glaucomatous locations and not others.

Fellman et al. (1988) reported that, when establishing threshold with a larger stimulus area, a greater reduction in threshold luminance from that of a smaller stimulus area

was found in those with glaucoma compared to normal controls, likely due to the greater spatial summation area in those with glaucoma. As this threshold difference was greater than that noted with a change in background intensity (while keeping stimulus area constant), they concluded that stimulus area had a greater influence on retinal sensitivity than contrast. They also noted that the opposite phenomenon occurred in normal control subjects, i.e. the change in background intensity had a greater influence on retinal sensitivity than the change in stimulus area, thus indicating a differing spatial summation between normal control subjects and those with glaucoma. Felius et al. (1997) investigated differences in spatial summation for different wavelengths of stimulus and background, and found the greatest difference in Ricco's area between normal control subjects and those with glaucoma with a white stimulus on a white background.

More recently, Redmond et al. (2010a) have investigated these findings further. In addition to finding an enlarged Ricco's area in subjects with glaucoma compared with normal controls, they noted that the spatial summation curve for glaucoma patients was displaced in a purely lateral direction, and could be laterally shifted to overlap exactly with that of healthy observers, such that sensitivity loss in glaucoma could be explained by stimulus area alone, rather than threshold luminance (*Figure 1.9*). This was true for both a white-on-white stimulus, and under S-cone isolation. Given that they also found that Ricco's area remains stable with age under photopic conditions (Redmond et al. 2010b), they proposed the potential for a direct clinical application for this finding.
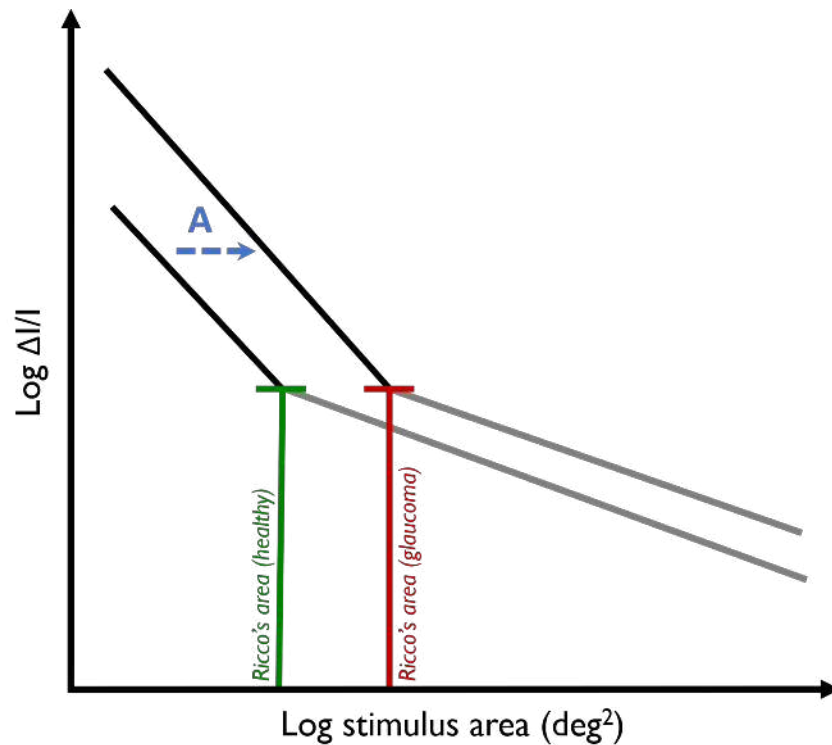
Figure 1.9 – Schematic illustrating the findings of Redmond et al. (2010a), in which a lateral shift of the summation curve, as indicated by 'A', was noted in the presence of glaucoma.

1.4.5   A potential perimetric application for spatial summation differences

Redmond et al. (2010a) proposed that this knowledge of a differing spatial summation curve in the presence of glaucoma could be utilised perimetrically. They noted that a stimulus which is area-scaled in accordance with changing spatial summation would result in a constant threshold, in contrast with a standard Goldmann III stimulus in which sensitivity would decline, dependent on its area relative to Ricco's area.

The series of studies presented in this thesis will attempt to exploit the difference in spatial summation in the presence of glaucoma, into a clinically useful perimetric test, by investigating stimuli that vary not only in contrast, as is currently utilised in SAP, but also in area.

The concept of a perimetric stimulus that is not of a fixed area has been previously investigated. Frisén (1987) used high-pass resolution ('vanishing') targets of varying size to determine threshold, and studies indicated that variability was more uniform across the range of sensitivity values than that of SAP (Chauhan and House 1991; Wall

et al. 1991). Wall (2013) investigated the potential of 'Size Threshold Perimetry' (STP), in which the area of the stimulus varied, while luminance remained constant, and threshold was determined as the smallest perceivable stimulus. Wall (2013) compared threshold measurements from STP with those of Goldmann III, V, and VI stimuli, and found a statistically different, but clinically similar, number of abnormal visual field locations with all testing strategies. However, it should be noted that the protocols for these four testing strategies were quite different, which may impede a direct comparison. The three Goldmann stimuli were presented using a Humphrey Field Analyzer, with the Goldmann III stimulus investigated using the SITA standard 24-2 algorithm, and Goldmann V and VI investigated using a Full Threshold strategy. The use of a thresholding algorithm, such as SITA, to establish sensitivity will yield a different threshold estimate than if a Full Threshold strategy is utilised (Artes et al. 2002a). Additionally, STP was presented on an entirely different test platform, with stimuli displayed on a computer-controlled, touch-sensitive monitor, utilising a background luminance of 15.9 cd/m^2 (5.9 cd/m^2 above the background luminance of a Humphrey Field Analyzer). The stimulus luminance was set at 25.5 cd/m^2, although it is unclear why this luminance value was selected. In addition, the observer indicated the location of the stimulus they had seen on the monitor, requiring a greater level of accuracy than that of the Humphrey Field Analyzer, in which a button is pressed when a stimulus is perceived, but the stimulus location is not identified by the observer. Perhaps the most crucial difference between these four testing strategies is the difference between the stimulus scales. Despite all utilising a dB scale, energy increments are not equal for stimuli of differing area, as described in section *1.3.2.3* for Goldmann stimuli, impeding a direct comparison between these stimuli. Additionally, as the dB scale employed by the STP strategy uses area increments rather than luminance increments, it cannot be considered the same measurement scale. As comparisons conducted by this study were therefore not direct between stimuli, conclusions drawn are likely to be subject to bias.

The series of studies presented in this thesis differs from those conducted previously, in that stimuli have been designed with the express purpose of identifying spatial summation differences between healthy subjects and those with glaucoma, which was

not a consideration in the area-modulating strategies discussed above. In addition, the stimuli are not restricted to the five Goldmann stimuli that have dominated perimetric studies for much of the past 70 years. In doing so, the study aim is to address some of the limitations of the currently utilised SAP, as outlined in section *1.3.3*.

1.5 Summary and study outline

This first chapter has provided a brief introduction to POAG, i.e. the disease investigated in this series of experiments, the research to date on perimetric techniques that have attempted to identify changes in the visual field due to cell damage caused by this disease, and the concept of spatial summation and findings which suggest its potential in the design of perimetric stimuli, which may more readily target the reorganisation that occurs in response to glaucomatous damage.

The series of experiments presented here compares four stimulus forms. One stimulus (A) varies in area, while maintaining a constant contrast, and is designed such that the area modulations begin within the area of complete spatial summation. One stimulus (AC) varies in both area and contrast simultaneously, modulating from a small, dim stimulus, to a larger, brighter stimulus; again, it is designed such that the modulations begin within the area of complete spatial summation. One stimulus (C_R) varies in contrast only, but is smaller than the currently utilised Goldmann III stimulus, such that it falls within the area of complete spatial summation at all locations of the visual field. Finally, one stimulus (GIII) is equivalent to that of a Goldmann III stimulus, and serves as a control stimulus.

Prior to undertaking these experiments, the equipment set-up was carefully considered, and the characterisations necessary to understand how the choice of apparatus may impact the generation of stimuli has been detailed in chapter two. Preliminary investigations are described in chapter three. As a number of experiments utilise the method of constant stimuli (MOCS) strategy, this includes an exploration of the MOCS procedure, the construction of frequency-of-seeing (FOS) curves and the subsequent fitting of psychometric functions, and actions that have been taken to avoid some of the known biases that may occur with these procedures have been

described. Also detailed in this chapter is an experiment conducted to investigate response variability profiles of stimuli of differing areas.

The four stimulus forms were then compared over three experiments. The primary experiment utilises a MOCS procedure to investigate the total deviation, response variability, and resulting signal-to-noise ratio (SNR) achieved with the four stimulus forms in the presence of glaucoma (chapter four). The subsequent experiment investigates the test-retest variability of these four stimulus forms (chapter five), and the final experiment investigates the robustness of these stimuli to optical imperfections (chapter six).

The final chapter presented here (chapter seven) considers the overall findings from the study, and suggestions for further research.

Chapter 2 Selection and characterisation of an appropriate display screen

2.1 Introduction

Conducting research using a commercially available perimeter can be very advantageous. A clinical set-up can be mimicked, such that the test is comparable to perimetry conducted in practice, and the instrument and test design will likely be familiar to participants from routine eye examinations, or ophthalmology appointments. However, they can also be somewhat limiting in their use for research purposes as they present certain restrictions. For example, in a projection perimeter, the range of stimulus areas that can be displayed are determined by an aperture wheel mounted inside the instrument, through which light is projected. In the Octopus Perimeter and Humphrey Field Analyzer, these stimulus sizes are fixed and range between Goldmann I, with an area of -2.02 log deg^2, and Goldmann V, with an area of 0.37 log deg^2 (Haag-Streit 2014; Zeiss 2014; Zeiss 2015). Within this thesis, a number of different stimulus parameters will be compared to determine what characteristics yield the optimal disease-signal to measurement-variability ratio for the detection and monitoring of glaucoma. The stimuli to be compared are:

- **C$_R$ stimulus**: This will vary in contrast while area (-1.92 log deg^2) remains constant. The stimulus area selected will be smaller than the standard Goldmann III stimulus commonly used in SAP, and will fall within the area of complete spatial summation.
- **GIII stimulus**: This will also vary in contrast while area remains constant, with an equivalent area to a Goldmann III stimulus (-0.95 log deg^2), as is commonly used in SAP.
- **A stimulus**: This will vary in area while contrast (0.70 log ΔI/I) remains constant.
- **AC stimulus**: This will vary in both area and contrast simultaneously.

For each stimulus parameter, the energy in each step-modulation (whether area, contrast or both) should be equivalent, according to *Equation 1.1*. Full details of the four stimulus forms are provided in chapter four.

If using a commercially available projection perimeter, the study would be limited by the five available stimulus sizes. The limitation this would present is best illustrated by examining the AC stimulus, which varies proportionally in both area and contrast. As each Goldmann stimulus increases in area by a factor of four (i.e. the area of each Goldmann stimulus is four times larger than the previous Goldmann stimulus), the luminance would also be required to increase by a factor of four with each step. If the example of a 0.15 deg² stimulus (Goldmann III) is considered at a luminance of 100 cd/m², and a standard duration of 0.2 seconds, this equates to an energy of 3.0 cd/m^2.s.deg^2 (*Equation 1.1*); this energy is equivalent to that of a Goldmann III stimulus of 15.03 dB on the Humphrey Field Analyzer. If this is then varied as an AC stimulus, the next step must be 0.58 deg² (Goldmann IV) at a luminance of 400 cd/m², equating to an energy of 46.4 cd/m^2.s.deg^2; this energy is equivalent to that of a Goldmann III stimulus of 3.13 dB. These step-modulations are simply too large to accurately measure sensitivity, and subsequently response variability, as required.

In view of such technical limitations with existing commercial perimeters, the ViSaGe MKII visual stimulus generator (Cambridge Research Systems, Rochester, UK), in combination with a display screen, was selected to run the experiments presented within this thesis. This equipment allows greater control and flexibility over stimulus generation and display than commercially available projection perimeters. Crucially for this particular series of experiments, it permits greater freedom of choice regarding stimulus area, and is not restricted to the standard Goldmann stimuli. However, as with all experimental set-ups, it is not without limitation itself. As psychophysics is defined as 'a scientific discipline to determine the relationship between a physical stimulus and a perceptual response' (Comerford et al. 2002), it is important to fully understand the physical stimulus to ensure a full understanding of the perceptual responses from observers. As such, detailed knowledge of the experimental set-up is required, including any artefacts that may result from image generation on the class of display screen used.

2.2 Display Screen Options

In any computer-based system, images are presented as a series of discrete elements (pixels), each of which emits a particular luminance and colour. On a chromatic display, the outcome colour of a pixel is determined by the additive mix of three monochromatic colours (red, green and blue). When displayed simultaneously and adjacently, the individual pixels appear as one continuous image. Different types of display screen generate and display pixels using different methods.

An ideal display screen for these experiments would have a high resolution (thus allowing fine and accurate increments in stimulus area), and a high maximum luminance (ensuring an adequate dynamic range in contrast-modulating stimuli, to permit threshold measurement in those with more advanced visual field loss). In addition, low variability in luminance output, both intra- and inter-test, and consistency of both stimulus dimensions and luminance output across the display screen (spatial homogeneity) would be required. Cathode Ray Tube (CRT) displays are the traditional choice of display screen for psychophysical, vision-based research (Metha et al. 1993; Bach et al. 1997; Krantz 2000; Brainard et al. 2002), but they are becoming increasingly difficult to obtain, and have been largely replaced in the commercial market by Liquid Crystal Displays (LCDs), and more recently by Organic Light Emitting Diode (OLED) displays. All display types have advantages and disadvantages, which are briefly discussed here.

2.2.1 CRT

As CRT displays have been the traditional choice for vision-based research for many years, their advantages and disadvantages are well documented (Metha et al. 1993; Bach et al. 1997; Krantz 2000; Brainard et al. 2002). A schematic diagram of a chromatic CRT display is shown in *Figure 2.1*. Held within a vacuum tube, heated cathodes from three guns (one each associated with the red, green and blue components of each pixel), emit a stream of negatively charged electrons towards the front, glass screen (faceplate). These are focused, accelerated and deflected by positively charged magnetic and/or electrostatic electrodes (anodes). The electron beams travel in different directions towards the shadow mask, a screen near the faceplate containing numerous apertures, and are either deflected away from the

shadow mask or pass through the apertures to hit the faceplate. A layer of phosphor coats the back of the faceplate, distributed in discrete areas that radiate a pulse of red, green or blue light when stimulated by the electrons. The structure of the electrodes and shadow mask are such that electrons from the red, green and blue guns are directed towards red, green and blue emitting phosphors, respectively. The input from the computer is transformed into three voltage signals, which adjust the output power of each of the three electron guns, and thus the luminance of each of the three phosphor elements. The additive mix of these three colour voltages determines the perceived colour and luminance for each pixel. (Metha et al. 1993; Cowan 1995; Mollon 1998; Robson 1998).
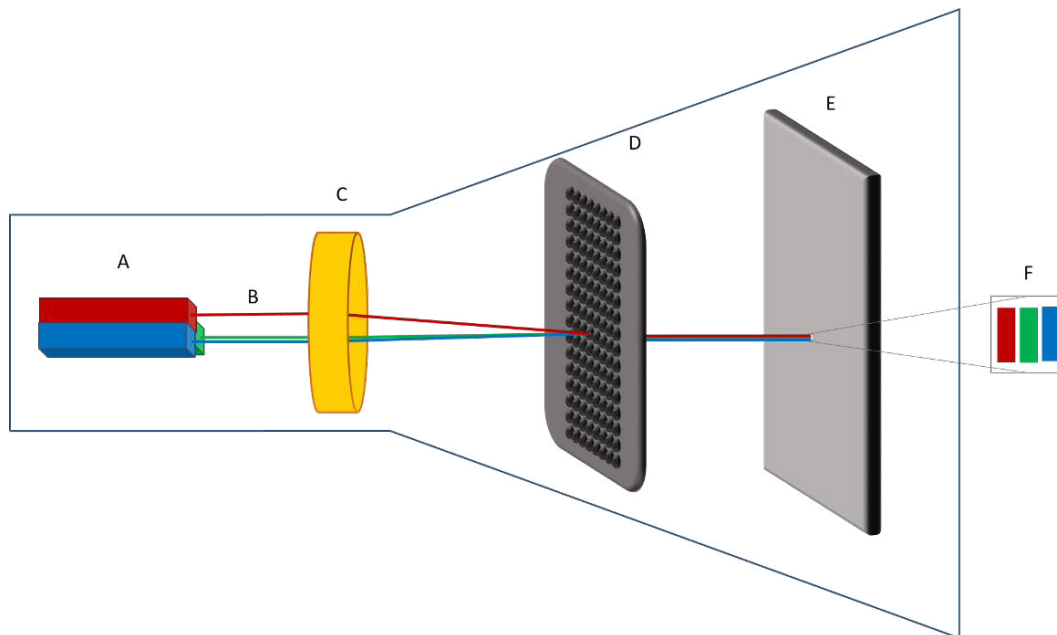
Figure 2.1 – Schematic diagram of a CRT display. Following an input signal, electron guns (cathodes; A) emit electron beams (B). These beams are focused, accelerated and deflected by magnetic and/or electrostatic electrodes (anodes; C) towards the shadow mask (D). Here, they are either deflected away or pass through the apertures to hit their respective phosphors on the faceplate (E), an enlargement of which is shown in (F). Adapted from Mulholland (2014), Cowan (1995) and Metha et al. (1993).

2.2.1.1 Spatial characteristics

The resolution of a CRT display is not fixed, and may be adjusted within a certain range, although an increase in resolution is often accompanied by a reduction in

refresh rate. In addition, pixels often cannot be considered completely independent, with some luminance 'bleed' into adjacent pixel space due to the scanning nature of the electron beams (Figure 2.2).
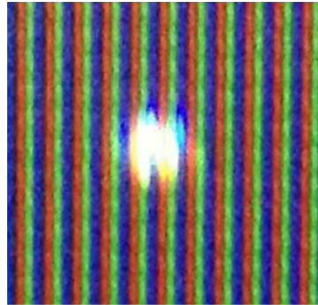
Figure 2.2 – Single pixel presented on a CRT display, showing pixel bleed into adjacent pixel areas.

2.2.1.2 Temporal characteristics

The pulse of light emitted by the phosphor on excitation by the electron beam is very short lived, rising to a peak and then decaying away to approximately 10-30% of the peak, depending on the phosphor type (Sperling 1971; Elze 2010a), as illustrated in *Figure 2.3*. This typically occurs over less than ten milliseconds, with a duty cycle (the ratio of the ON-period to the desired duration) of between 8 and 25% (Zele and Vingrys 2005; Elze 2010b). This can pose difficulties when attempting to specify an accurate stimulus duration, although, as the CRT has been shown to maintain accurate luminance profiles, with little interaction between consecutive frames, duration will be consistent between stimulus presentations (Bridgeman 1998; Ghodrati et al. 2015; Mulholland et al. 2015d).

The scanning pattern of the electron beam is known as a raster scan (*Figure 2.4*), which must occur repetitively and systematically to maintain a constant image on the screen, due to the phosphor decay. A frame is defined as the time taken for a complete set of scans from the upper left corner of the display screen to the lower right corner, and the refresh rate denotes how many complete frames occur in one second (Cowan 1995; Mollon 1998; Robson 1998), measured in hertz (Hz). The frame rate is the

number of frames per second the display can receive from the computer it is connected to, measured in frames per second (FPS).
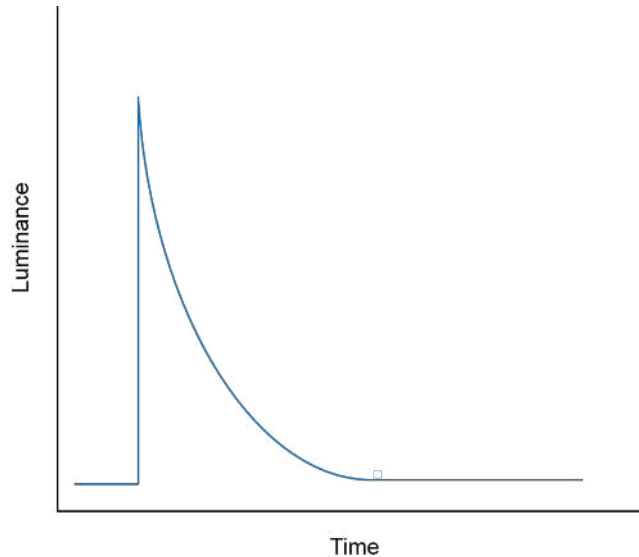
Figure 2.3 – Schematic of a phosphor activation and subsequent exponential decay in a CRT display.
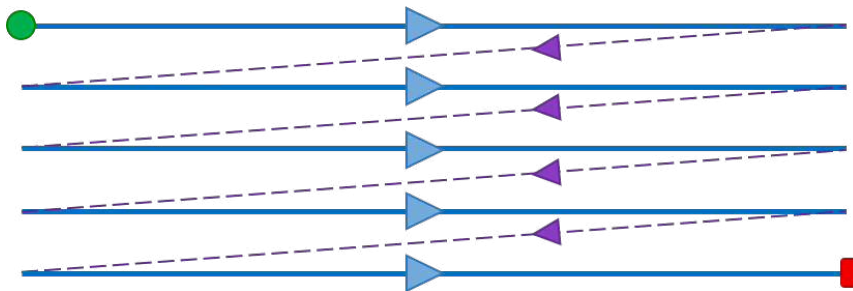
Figure 2.4 – Typical raster scan pattern of a CRT display. The electron beam starts at the upper left corner of the display screen (green circle) and scans from left to right (blue, solid lines and arrows). The beam returns to the left side of the screen (purple, dashed lines and arrows) before scanning the next line. Once it has reached the lower right corner of the screen (red rectangle), it begins again at the upper left corner (green circle).

Refresh rate is unfixed in the majority of CRT displays, and can be adjusted as required. As the visual system can perceive flicker at low frequencies, the recommended refresh rate should be higher than the 'critical flicker frequency' (CFF), with 60 Hz often

selected to ensure that a constant, steady image is perceived across the visual field. Despite the perception of a constant image, neural activity in retinal ganglion cells and artefacts in cortical cells have been reported with the use of frame rates below 100 Hz (Shady et al. 2004; Zele and Vingrys 2005), resulting in the suggestion that a refresh rate higher than this should be used in vision-based experiments. However, an increase in the spatial resolution of a CRT display is generally accompanied by a decrease in refresh rate, and a compromise between these two factors must be reached.

2.2.1.3 Luminance output

The CRT display is known to suffer from spatial inhomogeneity, such that a command given to the computer to produce a uniform background luminance may not be measured as such across the entirety of the display. Luminance output is often reported as highest in the central area for CRT displays, reducing towards the periphery, and may vary by as much as 27% (Metha et al. 1993; Krantz 2000). This occurs due to the angle of the electron beams through the shadow mask.

A significant warm-up period is also required to permit thermal equilibrium of the cathodes, and stabilise the luminance output. This varies substantially between CRT displays, reported as between 40 and 150 minutes, after which fluctuations of ~1% should still be expected (Metha et al. 1993; Krantz 2000; Brainard et al. 2002; Klein et al. 2013).

2.2.2   LCD

In an LCD, light is emitted from a source at the back of the display, and passes through three layers: the first polarising layer, a layer of liquid crystals, and a second polarising layer, orientated at 90° to the first polarising layer. Voltage passing through the liquid crystals determines their alignment, and therefore how much light can pass through the filters.

LCDs have several advantages over CRT displays, for example they take up less space due to the flat screen, and they have better pixel independence. However, for vision-based research there are some important disadvantages. Ghodrati et al. (2015) reported that measured luminance is greatly reduced with viewing angle, up to 80% at

45° from fixation in some instances; this was contrary to that found with a CRT display, in which measured luminance was almost independent of viewing angle. This is in addition to a spatial inhomogeneity similar to CRT displays; i.e. the brightest luminance output is often reported at the centre of the display, with a reduction towards the periphery.

The change in luminance from one image to another is determined by the speed of the liquid crystal alignment (response time). This has been demonstrated as highly variable, both between different monitor types, and within the same monitor with different levels of luminance (Elze 2010b; Ghodrati et al. 2015). The same studies showed that LCDs could not reach maximum luminance within a single frame, and there was a substantial lag in luminance change, such that a black frame following a white frame could appear as grey.

As this series of experiments is concerned with peripheral threshold and response variability, precise timings are required between presented stimuli. The disadvantages highlighted here therefore render the LCD unsuitable for this series of experiments.

2.2.3 OLED displays

OLED displays are an emerging technology in vision science experiments, therefore not as many psychophysical studies have used an OLED display as part of their experimental apparatus. In an OLED display, as shown in *Figure 2.5*, a thin film of organic, electroluminescent material is located between two electrodes, the transparent anode and the reflective, metal cathode, all deposited on a substrate. When a voltage is applied, electrons are injected from the cathode into the Lowest Unoccupied Molecular Orbital (LUMO) of the organic material, and holes are injected from the anode into the Highest Occupied Molecular Orbital (HUMO) of the organic material. These charges recombine in the organic layer to form an exciton, an electrostatically bound electron-hole, the decay of which results in energy being emitted radiatively as a photon. Different organic materials result in either red, green or blue emissions. The overall output colour of the pixel is determined by the mix of these three, and the brightness determined by the voltage (Davidson-Hall et al. 2017;

Geffroy et al. 2017). Alternatively, different colour filters can be placed in front of a single diode (Farrell et al. 2017).
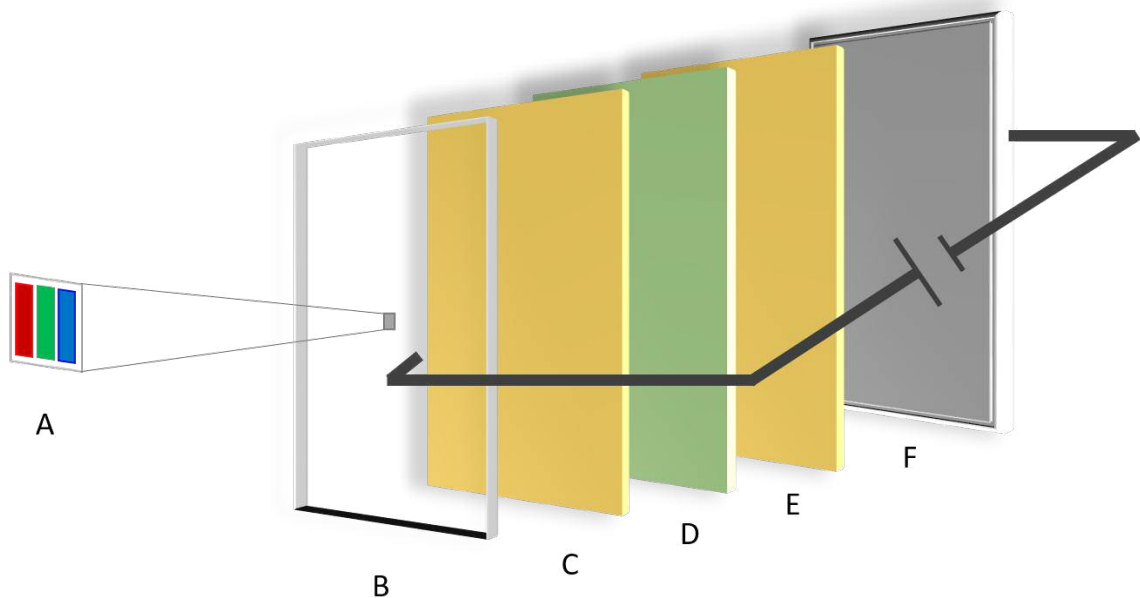
Figure 2.5 – Schematic diagram of an OLED display. A voltage is applied between the metallic cathode (F) and the transparent anode (B), causing electrons to be injected in the Lowest Unoccupied Molecular Orbital (E), and holes to be injected into the Highest Occupied Molecular Orbital (C). These charges recombine in the organic layer (D) to form excitons, whose decay releases energy as a photon. Each pixel is made up of red, green and blue subpixels (A); the different colours are produced by differing organic materials in (D). Adapted from Geffroy et al. (2017) and Davidson-Hall et al. (2017).

2.2.3.1 Spatial characteristics

An OLED display has a fixed spatial resolution which cannot be adjusted, unlike the CRT display. Although the maximum resolution of a CRT may be higher than that of an OLED, the use of this resolution may only be used in combination with a substantially reduced refresh rate, often to unacceptably low levels for vision-based research (Zele and Vingrys 2005); an OLED display often achieves a higher resolution than that of a CRT display, while maintaining an acceptable refresh rate.

As pixels are controlled individually in an OLED display, each pixel is discrete, with an absence of the 'bleed' observed in a CRT display, as shown in *Figure 2.6*.
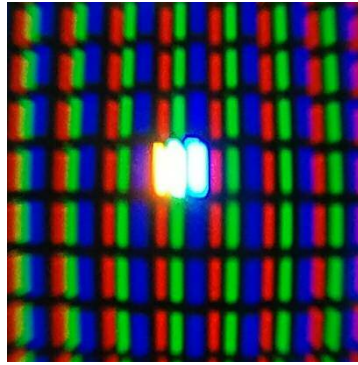
Figure 2.6 – A single pixel, presented on an OLED display, demonstrating the discrete nature of the pixels. Any apparent pixel bleed is due to the optics of the system used to take the photograph, and not due to the OLED display itself.

2.2.3.2 Temporal characteristics

In an OLED display, photon emission occurs as a pulse, with an onset, a period of fixed luminance output, and an offset. The luminance profile of a photon in an OLED display is shown in the schematic of *Figure 2.7*. A sharper onset and offset occurs in an OLED display than is seen with a CRT or an LCD, and the response time is much faster and more consistent than that observed with an LCD, remaining independent of the luminance output in the previous frame (Elze et al. 2013). The duty cycle has been reported as between 40 and 66% (Cooper et al. 2013; Elze et al. 2013), substantially higher than that of a CRT dislay (Zele and Vingrys 2005; Elze 2010a).

The refresh rate of an OLED display is often fixed at 60 Hz; this may be a disadvantage given the findings in refresh rates < 100 Hz, and the subsequent recommendations for vision-based experiments as noted in section *2.2.1* (Shady et al. 2004; Zele and Vingrys 2005).
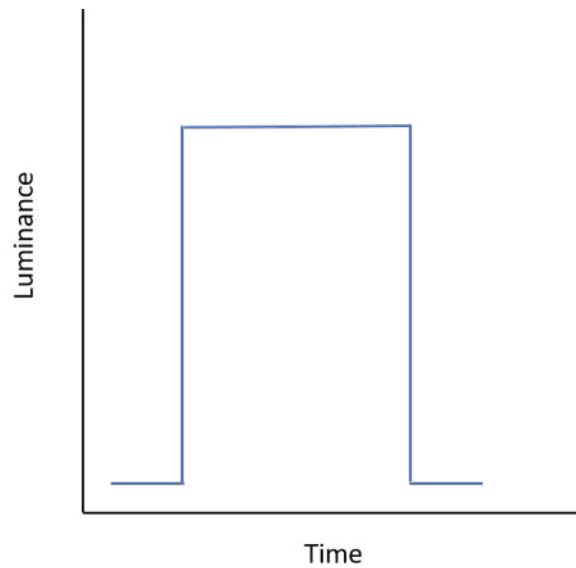
Figure 2.7 – Photometric output for an OLED display, showing sharp on- and off-sets.

2.2.3.3 Luminance output

As pixels are controlled individually in an OLED display, spatial inhomogeneity is markedly reduced compared to CRT displays. Ito et al. (2013), evaluating the characteristics of an OLED display, noted only an 8% reduction in luminance output towards the peripheral edge of the screen, and found that the OLED exhibited a truer black, measured as near 0 cd/m^2. They also noted a short warm-up time of just 15 minutes. They did, however, find a change in the luminance output under certain conditions. When displaying a bright target with an area greater than 40% of the screen, an overall reduction in the luminance output was measured. Equally, if a bright background were displayed, a reduction in target luminance was measured. This phenomenon appeared to be due to in-built luminance control characteristics of the particular OLED display used in the study, although has also been observed by Elze et al. (2013) with a different model of OLED display. This is not considered a disadvantage of OLED displays for this series of experiments, given that a low background of 10 cd/m^2 will be used, and no perimetric stimulus will be larger than 40% of the total display area.

Luminance was noted to decrease with viewing angles greater than 10°, suggesting a luminance reduction of ~13% at 30° from fixation for a white stimulus. Luminance

appeared to decrease more quickly with the blue than the red or green output, resulting in a white target appearing bright cyan with viewing angles greater than 80° (Ito et al. 2013). However, the authors acknowledged that the manufacturing company were aware of this colour shift and had counteracted it in later models. As this series of experiments is concerned with a maximum viewing angle of 30°, given that this is the maximum eccentricity commonly tested in SAP, this is not considered a disadvantage of the display for this series of experiments.

One of the main disadvantages at present is cost, as OLED displays are significantly more expensive than either CRT displays or LCDs.

Given the advantages and disadvantages of these types of display, the LCD was deemed unsuitable. While some generalisations may be made regarding display type, the individual display used for research should be fully characterised to ensure that limitations are identified and understood. As such, a 22" Hewlett Packard p1230 chromatic CRT display, and a 25" Sony PVM-A250 Trimaster El OLED display, were characterised to determine which was the most appropriate to use in this series of experiments, as detailed in section *2.3*.

2.3   Display characterisation

The 22" Hewlett Packard p1230 chromatic CRT display was set and maintained at a spatial resolution of 1024 x 768 pixels, with a refresh rate of 120 Hz. Brightness was fixed at 100%, and contrast at 70%, permitting a maximum luminance output of ~109 cd/m^2.

The 25" Sony PVM-A250 Trimaster El OLED display has a fixed spatial resolution of 1920 x 1080 pixels, and a refresh rate of 60 Hz. This OLED display has the option of a 'flicker-free' mode; when this is engaged, the same frame is presented twice within the same time-span as one frame in the absence of the flicker-free mode. This results in the display operating at 120 Hz, although the content is still updated from the computer at 60 Hz (Cooper et al. 2013; Elze et al. 2013), thus meeting the recommendations for vision-based experiments (Shady et al. 2004; Zele and Vingrys 2005).

Due to some uncertainty regarding the lifespan of the OLED, and in particular the possibility of reported colour imbalance due to unequal degradation of the three colour components (Cooper et al. 2013), the display was set and maintained at 70% brightness, and 80% contrast. This permitted a maximum luminance output of ~208 cd/m^2.

These settings were kept constant throughout all the experiments presented in this chapter.

2.3.1   Gamma correction

Display screens have a non-linear transfer function, such that the output luminance is not linearly related to the input voltage, leading to the gamma relationship described in *Equation 2.1.*

$$\phi(E) = \beta E^\gamma$$

Equation 2.1
(Metha et al. 1993)

$\phi(E) = Luminance\ output$

$E = Applied\ voltage$

$\beta\ = Constant$

$\gamma = Exponent\ of\ the\ power\ function$

This is a known consideration in CRT displays, as different electron guns have different β and γ values. As this can lead to distortions in the luminance and colour of the image displayed, this non-linearity must be corrected before commencement of any vision-based research, a process termed 'gamma correction'. This is achieved by taking various measurements across the full range of luminance outputs, with white light and with each of the individual colour components (red, green and blue). A function is

then fitted to these data (Robson 1998). Regular gamma correction is deemed necessary with CRT displays, as characteristics reportedly changed over short periods of time; at least once a month is usually recommended (Metha et al. 1993; Cowan 1995). Although an OLED display is generally considered to be much more stable than a CRT display, a gamma correction is still appropriate to ensure a linear relationship between voltage and luminance output, although it is unclear from the literature how often this should be repeated.

A gamma correction of both the Hewlett Packard p1230 chromatic CRT display, and the Sony PVM-A250 Trimaster El OLED display, was achieved using a ViSaGe MKII stimulus generator and ColorCAL II colorimeter (Cambridge Research Systems, Rochester, UK), with accompanying MATLAB code. An example of the resulting gamma correction curves from the Sony OLED display is shown in *Figure 2.8*.

Once gamma compensation has been applied, the voltage-luminance relationship can be expected to be fairly linear. This process was conducted prior to commencing any experiments presented in this chapter, and was repeated monthly in the selected display while further experiments (as described in chapters four to six) were ongoing.
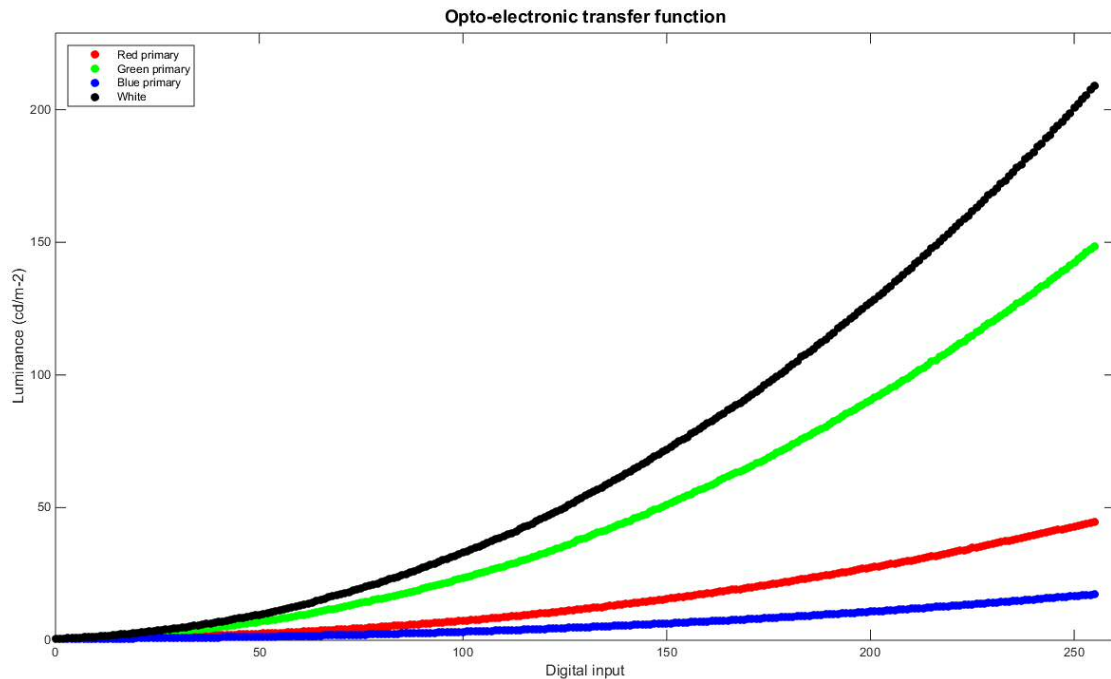
Figure 2.8 – Example gamma correction curves, obtained from a Sony PVM-A250 Trimaster El OLED display, using the ViSaGe MKII stimulus generator and ColorCAL II.

A series of experiments was conducted to characterise each display. One measured the time taken for luminance output to stabilise from the moment the display was first switched on, one investigated the spatial inhomogeneity of the luminance output across the display, both short-term (over a period of approximately 135 minutes) and longer-term (on three separate days, a maximum of 19 days apart), and one investigated whether stimulus area differed across the display.

2.3.2   Experiment one - luminance output stabilisation

2.3.2.1 Method

A (nominally) uniform, achromatic background of luminance ~10 cd/m², and chromaticity co-ordinates of x = 0.311 and y = 0.326, was generated using the ViSaGe MKII stimulus generator (Cambridge Research Systems, Rochester, UK). This was the proposed background luminance for the series of experiments presented in this thesis, in keeping with the background luminance employed by many commercially available perimeters (Haag-Streit 2014; Zeiss 2014; Zeiss 2015). Experiments were programmed in MATLAB (version 2014b; The Mathworks, Inc., Natick, MA) using the CRS toolbox

(version 1.27, Cambridge Research Systems, Rochester, UK). A calibrated ColorCAL II (Cambridge Research Systems, Rochester, UK) colorimeter was used to measure the luminance output at the centre of the display. A black, circular outline was presented to demarcate the centre of the display, of diameter 57 mm (2551.8 mm^2), within which the ColorCAL II was aligned and placed flush against the screen to block any stray light (*Figure 2.9*). All external light sources were extinguished. Measurements were taken over an area of 254.47 mm² (18 mm diameter).

The ColorCAL II was immobilised on a stand to ensure that all measurements were taken from the same area of the display. A MATLAB program was constructed to record the luminance output at approximately three-second intervals over a period of approximately 60 minutes. Measurements commenced from the time the monitor was first switched on and initialised (< 10 seconds for each display).

Figure 2.9 – Apparatus set-up to measure luminance output over time. The Sony PVM-A250 Trimaster El OLED display with a (nominally) uniform background of ~10 cd/m². A black circle demarcates the area in which the ColorCAL II was positioned to measure the luminance output. The ColorCAL II was immobilised on a stand to ensure that all luminance output measurements were taken at the same location. The same process was also conducted with the Hewlett Packard p1230 chromatic CRT display.

A further experiment, whereby luminance values were recorded at approximately 1-minute intervals over a period of approximately 360 minutes (6 hours), was conducted on a separate day, to determine the stability of the luminance output on days when multiple sets of data would be collected.

These experiments were conducted with both the Hewlett Packard p1230 chromatic CRT display, and the Sony PVM-A250 Trimaster El OLED display.

2.3.2.2 Results

Results are shown for the experiment conducted over 60 minutes in *Figure 2.10*. The light grey solid line indicates a luminance output of 10 cd/m^2, the desired luminance output. The light grey dashed lines indicate 10 cd/m^2 ±1%, in accordance with the expected luminance stability (Metha et al. 1993), and the light purple arrow indicates the time at which luminance output first fell within these parameters.

As an exact background luminance can be quite difficult to achieve, the mean luminance output of each display was also determined. This was established as the mean of all luminance measurements beyond 9.9 cd/m^2 (10 cd/m^2 -1%), and are shown on each plot. The dark purple arrow indicates the time at which luminance output first fell within mean luminance -1%. As mean luminance was substantially higher than 10 cd/m^2 for the CRT display, this is indicated by the dark grey, solid line, and ±1% values are indicated by the dark grey, dashed lines (*Figure 2.10.A*).

From this, it is apparent that the OLED display stabilised quickly (*Figure 2.10.B*), with luminance within 1% of the intended output of 10 cd/m^2 at 8.53 minutes. Although there were small fluctuations in the luminance output, these fluctuations were always within ±1% of 10 cd/m^2, although there appears to be a subtle increase in luminance output over the time period in which measurements were taken. Mean of all luminance measurements beyond 9.9 cd/m^2 was 10.02 cd/m^2, and luminance output was first within -1% of this value at 10.20 minutes.

The CRT display took slightly longer than the OLED, with luminance output within -1% of the intended output of 10 cd/m^2 at 11.97 minutes. However, beyond this time luminance output continued to increase. Mean of all luminance measurements

beyond 9.9 cd/m^2 was 10.66 cd/m^2, and luminance output was first within -1% of this value at 23.15 minutes. However, this does not appear to be truly representative of luminance stability, as luminance output did not remain within ±1% of 10.66 cd/m^2, and continued to increase throughout the 60 minutes.
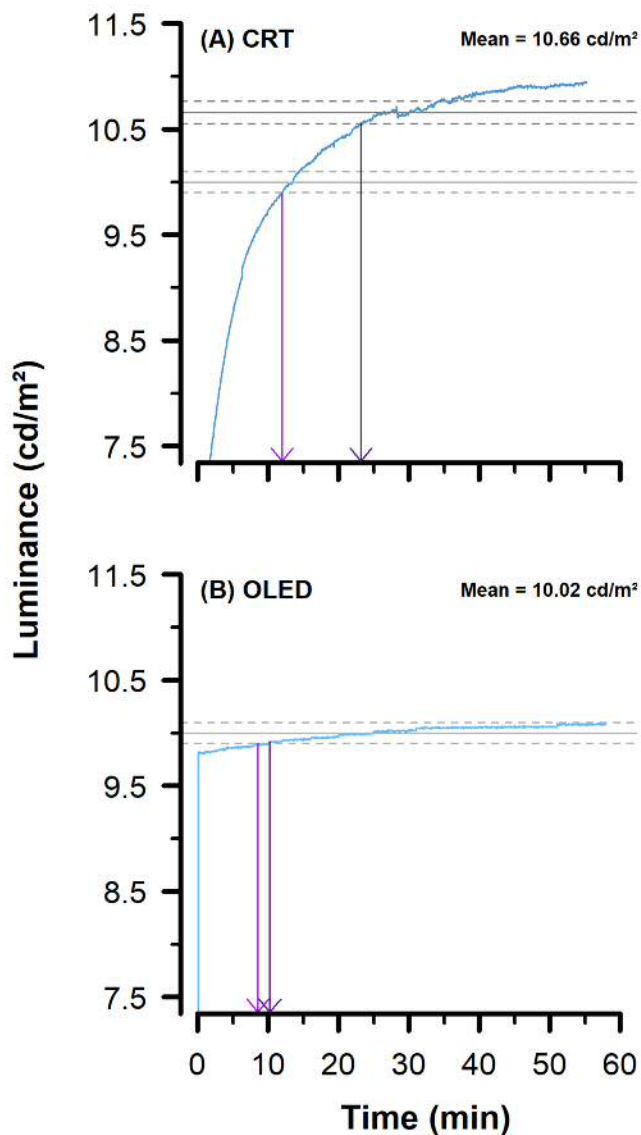
Figure 2.10 – Luminance output of the (A) Hewlett Packard p1230 chromatic CRT display, and (B) Sony PVM-A250 Trimaster El OLED display (cd/m²) with time (minutes) over a period of approximately 60 minutes. The first reading was denoted as '0' minutes, and readings were taken every three seconds. The light grey, solid line indicates 10 cd/m^2 (the intended background luminance). The light grey, dashed lines indicate ±1% of 10 cd/m^2, and the light purple arrow indicates the time at which luminance was first within these parameters. The dark grey, solid line indicates the mean luminance output, and the dark grey, dashed lines indicate ±1% of this value (shown in (A) only). The dark purple arrow indicates the time at which luminance output was first within -1% of the mean luminance.

Results are shown for the experiment conducted over 360 minutes in *Figure 2.11*. With the CRT display, a small but sharp increase in luminance output was noted at 102.53 minutes, after which luminance output appeared to stabilise (*Figure 2.11.A*). Mean of the luminance measurements beyond 102.53 minutes was taken as the stabilised mean luminance output of the CRT display. For the OLED display (*Figure 2.11.B*), mean luminance was taken as 10.00 cd/m^2, the intended luminance output. The solid, grey lines indicate the mean luminance output of each display. The dashed grey lines indicate ±1% of the mean luminance, and the dotted, grey lines indicate ±0.5% of the mean luminance. The light purple arrow indicates the time at which luminance output first fell within -1%, and the dark purple arrow indicates the time at which luminance output first fell within -0.5% of the mean luminance for each display.

Over this time period, a truer stabilisation of the CRT display is observed. Luminance first fell within -1% of the mean luminance output (11.12 cd/m^2) at 102.53 minutes, and within -0.5% at 115.09 minutes. Luminance output remained within ±0.5% of the mean luminance output throughout the rest of the time tested.

With the OLED display, luminance first fell within -1% of the mean luminance at 16.42 minutes, and within -0.5% at 27.99 minutes. Luminance output remained within ±0.5% of 10.00 cd/m^2 throughout the rest of the time tested.
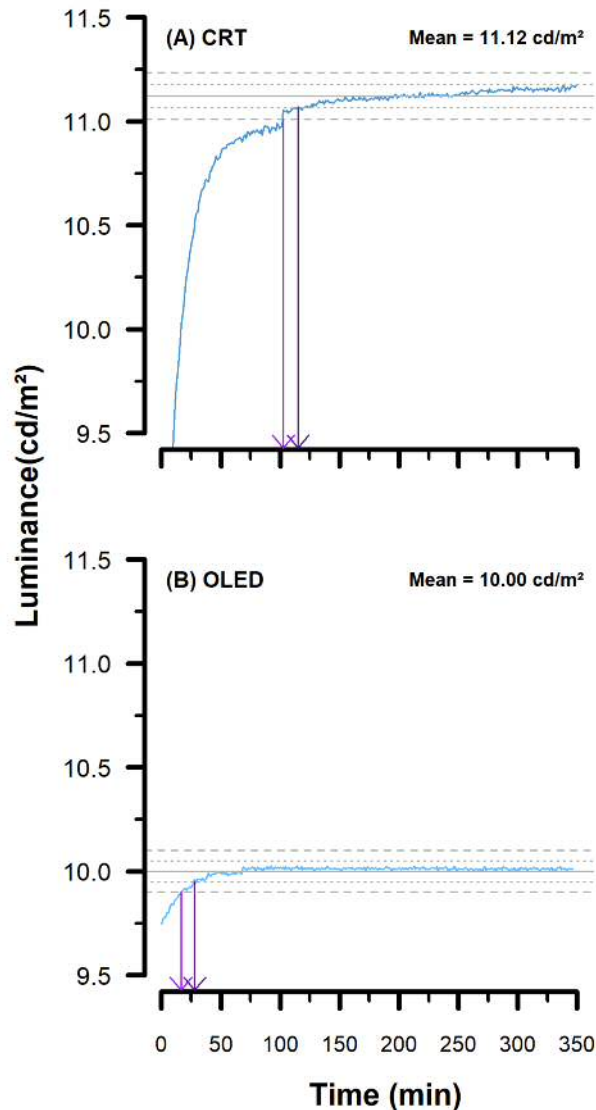
Figure 2.11 – Luminance output of the (A) Hewlett Packard p1230 chromatic CRT display, and (B) Sony PVM-A250 Trimaster El OLED display (cd/m²) with time (minutes) over a period of approximately 360 minutes. The first reading was denoted as '0' minutes. Readings were taken at 1-minute intervals. The solid grey line indicates mean luminance output. The grey, dashed lines indicate ±1%, and the grey, dotted lines indicate ±0.5% of the mean luminance output. The light purple arrow indicates the time at which luminance output was first within -1%, and the dark purple arrow indicates the time at which luminance output was first within 0.5% of the mean luminance output.

2.3.3 Experiment two - spatial inhomogeneity of luminance output

2.3.3.1 Methods

As for experiment one (section *2.3.2*), a (nominally) uniform, achromatic background of luminance ~10 cd/m², with chromaticity co-ordinates of x = 0.311 and y = 0.326, was generated. The display was subdivided into 99 sections, demarcated by black, circular outlines of diameter 34.2 mm (918.6 mm^2, shown in *Figure 2.12*). These circles

denoted the areas in which the ColorCAL II was placed, flush against the display to block all external stray light, to measure the luminance output. As in section *2.3.2,* the measurements themselves were taken over a much smaller area of 254.47 mm² (18 mm diameter) within each circle. All external light sources were extinguished.
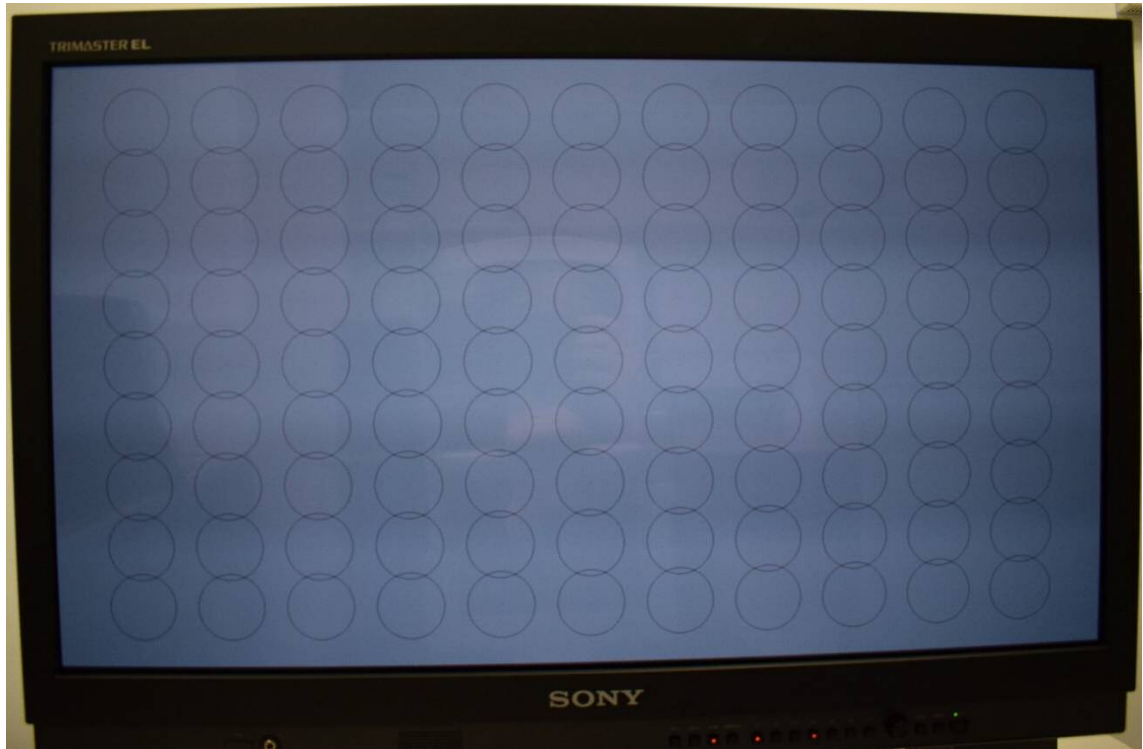
Figure 2.12 – Photograph showing the display on the Sony PVM-A250 Trimaster El, used to determine spatial inhomogeneity of the luminance output of the display. The black circles demarcated each area in which the ColorCAL II was positioned for measurement of luminance output. Tests were also conducted with the Hewlett Packard p1230 chromatic CRT display.

A calibrated ColorCAL II colorimeter was used to measure the luminance output at each of the locations shown in *Figure 2.12*. The ColorCAL II was manually positioned in the centre of each circle, and three consecutive luminance output readings were taken, over a time period of approximately 11 seconds. The mean of these three readings was then determined for analysis.

It took approximately 45 minutes to take a full set of readings, i.e. at all 99 locations. To determine the short-term fluctuation across the display, two further sets of readings were taken, with a time interval of approximately 45 minutes between

readings at each location. Day-to-day fluctuation was also determined by measuring the luminance output at each location on two subsequent days, within 19 days of the initial readings, giving luminance output measurements for three days in total. These were compared with the first set of readings on day one. These readings were used to determine the expected luminance output fluctuation throughout collection of different datasets (for experiments presented in chapters four to six), those occurring on the same day and those occurring on different days.

The CRT display was switched on approximately 115 minutes, and the OLED display was switched on approximately 30 minutes, prior to any measurements being taken, to allow adequate luminance stability, in accordance with the results obtained in section *2.3.2*.

2.3.3.2 Results

Luminance output measurements for the first set of readings at the 99 locations are shown in *Figure 2.13* for (A) the CRT display, and (B) the OLED display. The x- and y-axes denote the co-ordinates for each location (mm); a different scale is displayed along the x-axes in (A) and (B) as the two display screens differ in width, although are similar in height. Luminance output for each measured location is displayed, and luminance outputs between these locations have been predicted using bi-linear interpolation with MATLAB, and are indicated according to the colour scale on the right. The highest luminance output is denoted by the blue rectangle, and the lowest luminance output is denoted by the magenta rectangle for each display screen.
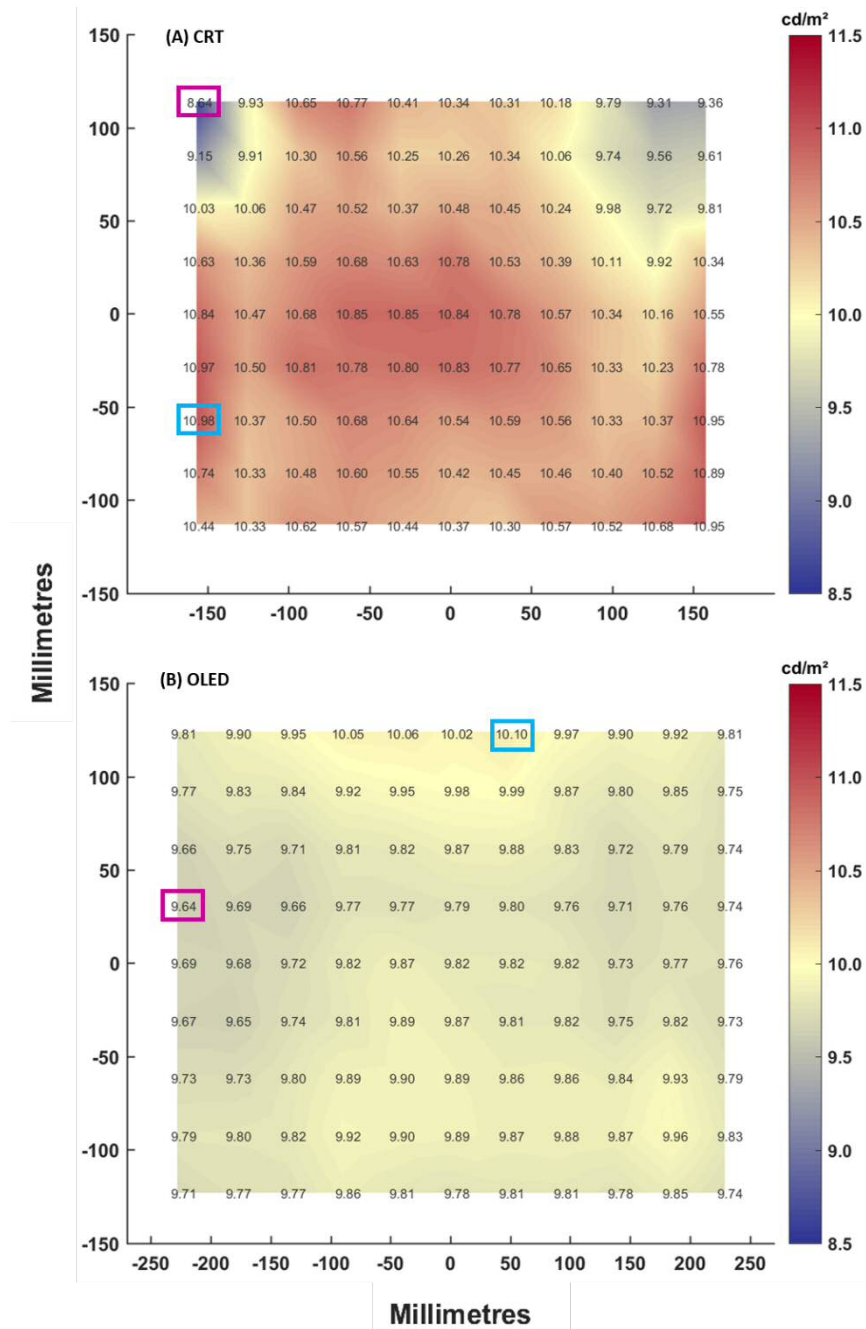
Figure 2.13 – Luminance output readings at 99 positions from the (A) Hewlett Packard p1230 chromatic CRT display, and (B) Sony PVM-A250 Trimaster El OLED display. The x- and y- axes denote the co-ordinates for each location (mm). The lowest luminance output is denoted by the magenta rectangle, and the highest luminance output is denoted by the blue rectangle.

As previously noted, studies have reported the highest luminance output at the centre of a CRT display (Metha et al. 1993; Krantz 2000). *Figure 2.13* shows that this did not hold true for either of the displays used here. While the centre of the CRT display was generally brighter, with a reduction in luminance output towards the periphery, the

extreme right and left lower areas had a higher luminance output than the centre. The highest luminance output was at a location near the lower left corner, and the lowest luminance output was at a location at the upper left corner. Mean luminance output was 10.39 cd/m^2, with a range of 2.34 cd/m^2 (8.64 to 10.98 cd/m^2).

The luminance output for the OLED display appeared more consistent, with no observed trend in the measurements. The highest luminance output was measured at a location at the top of the display, and the lowest luminance output was measured at a location to the left of the display. The mean luminance output was 9.82 cd/m², with a range of 0.46 cd/m² (9.82 to 10.10 cd/m²).

The range of non-uniformity was calculated for both displays using *Equation 2.2* and *Equation 2.3*.

$$LU = \frac{LM - ScM}{ScM} \times 100\%$$

<div align="right">*Equation 2.2*</div>

$LU = Luminance\ uniformity\ of\ given\ location$

$LM = Mean\ luminance\ output\ of\ 3\ consecutive\ readings\ at\ given\ location$

$ScM = Mean\ luminance\ output\ of\ all\ locations\ across\ display$

$$RNU = Max\ LU - Min\ LU$$

<div align="right">*Equation 2.3*</div>

$RNU = Range\ of\ non\text{-}uniformity$

The range of non-uniformity (RNU) was determined as 22.51% for the CRT display, and 4.72% for the OLED display.

Comparing all five sets of readings (three sets of readings on day one, and one set each on days two and three), the minimum RNU across the CRT display was 16.90%, and the maximum was 22.51%, a range of 5.61%. The minimum RNU across the OLED display was 4.29%, and the maximum was 5.01%, a range of 0.72%.

Variability over 135 minutes

Figure 2.14 shows the difference in luminance output with the CRT display with the (A) second and (B) third set of readings on the same day, compared with the luminance output from the first set of readings. *Figure 2.15* shows the same for the OLED display. As in *Figure 2.13*, the x- and y- axes denote the co-ordinates for each location (mm), and the difference in luminance output compared with the first set of readings are displayed for each measured location; a positive value denotes a higher luminance output than the first set of readings and a negative value denotes a lower luminance output. Luminance differences between these locations have again been predicted using bi-linear interpolation, and are indicated according to the colour scale on the right. Maximum luminance difference is denoted by the blue rectangle, and minimum luminance difference is denoted by the magenta rectangle.

For the CRT display (*Figure 2.14*), an increase in luminance output was noted at all locations. Locations at the peripheral areas of the display screen generally displayed greater variations than locations at the centre, particularly at the left side. The greatest increase in luminance output was observed at a location at the upper left corner for both the second and third sets of readings (0.31 and 0.50 cd/m^2 respectively), in comparison with the first, a maximum increase of 5.83%. The lower right corner appeared the most stable area, with the lowest increase in luminance output observed at a location in this area for both the second and third sets of readings (0.02 and 0.05 cd/m^2 respectively), although the exact location differed slightly from the second to the third readings.

For the OLED display (*Figure 2.15*), an increase in luminance output was noted at some locations, and a decrease at others; generally, an increase in luminance output was observed at the left side of the display, and a decrease in luminance output was observed at the right side, although this was more pronounced with the second set of

readings than the third. As with the CRT display, greater variations were generally observed at the left side of the OLED display, with the greatest difference in luminance output noted at a location in the upper left corner for both the second and third sets of readings (an increase of 0.081 and 0.102 cd/m^2 respectively) in comparison with the first, a maximum of 1.04%. The smallest difference in luminance output was observed at a location at the centre of the display screen with the second set of readings (an increase of 0.0002 cd/m^2), and at a location towards the lower right corner with the third set of readings (an increase of 0.0004 cd/m^2).
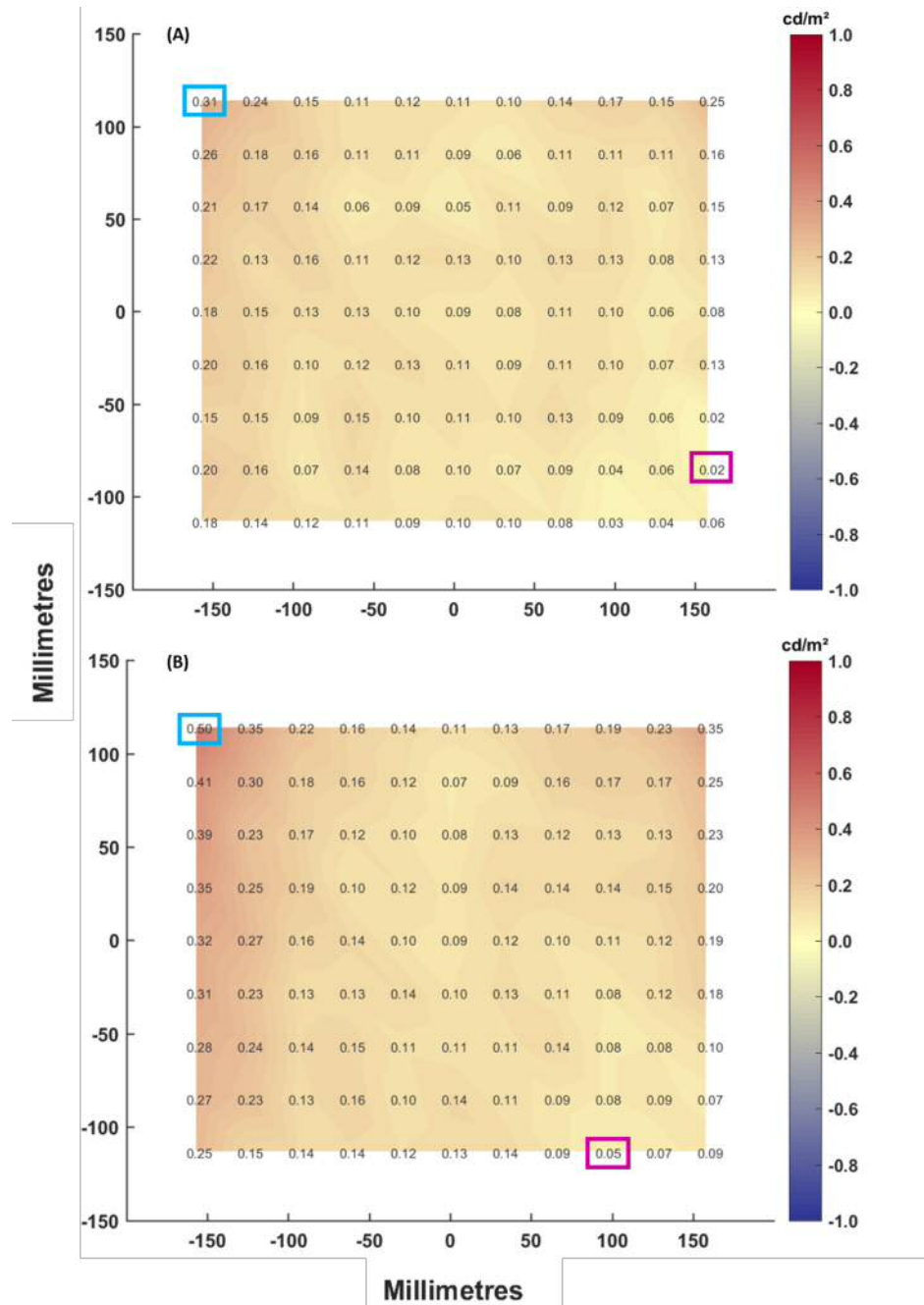
Figure 2.14 – Luminance difference in cd/m² of the (A) second and (B) third set of readings, compared to the first set of readings on day one with the Hewlett Packard p1230 chromatic CRT display. A positive value denotes a higher luminance output, and a negative value denotes a lower luminance output, compared with the first set of readings. The x- and y- axes denote the co-ordinates for each location (mm). Luminance differences have been interpolated between the measured locations, and are indicated according to the colour scale on the right. The maximum luminance difference is indicated by the blue rectangle, and the minimum luminance difference is indicated by the magenta rectangle.
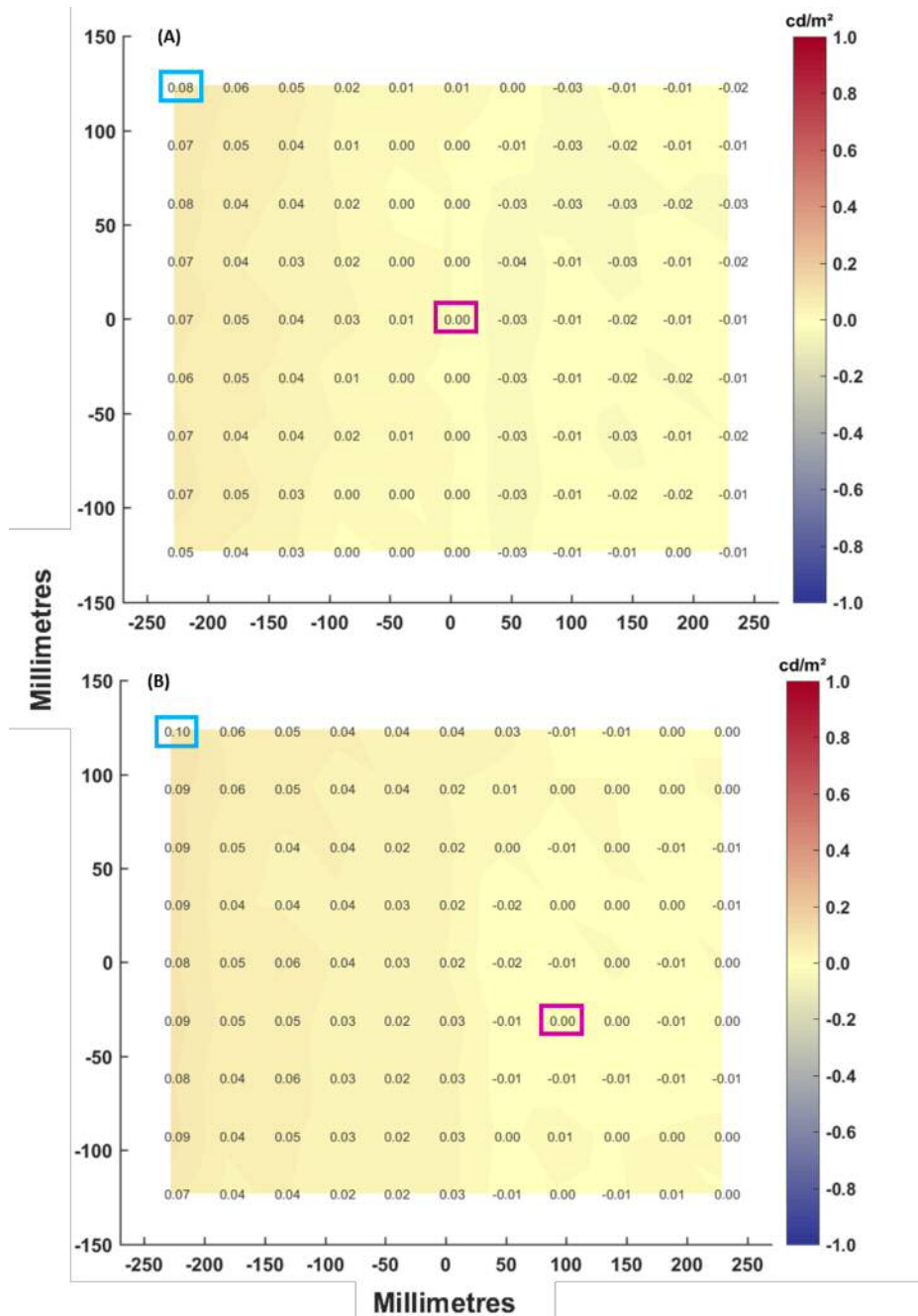
Figure 2.15 – Luminance difference in cd/m² of the (A) second and (B) third set of readings compared to the first set of readings on day one with the Sony PVM-A250 Trimaster El OLED display. A positive value denotes a higher luminance output, and a negative value denotes a lower luminance output, when compared with the first set of readings. The x- and y- axes denote the co-ordinates for each location (mm). Luminance difference has been interpolated between the measured locations, and are indicated according to the colour scale on the right. The maximum luminance difference is indicated by the blue rectangle, and the minimum luminance difference is indicated by the magenta rectangle.

Variability over three separate days

Figure 2.16 shows the difference in luminance output with the CRT display between the (A) second and (B) third day, in comparison with the first set of readings taken on day one. As previously, a positive value indicates a higher luminance output and a negative value indicates a lower luminance output when compared with the first set of readings.

Luminance output over the three days showed a greater variation in measurements than the short-term fluctuation over 135 minutes. Greater variations were observed at peripheral locations of the CRT monitor than at central locations. As previously, the greatest difference in luminance output was observed at a location at the upper leftmost corner (a greater luminance output by 0.68 cd/m^2 and 0.66 cd/m^2 on days two and three respectively, a maximum difference of 7.83%). The minimum variation was at slightly different locations on the two days, at a location near the top of the CRT display on day two (a lower luminance output by -0.003 cd/m^2), and at a location near the upper left corner on day three (a difference of 0 cd/m^2).

Figure 2.17 shows the difference in luminance output with the OLED display between the (A) second and (B) third day, in comparison with the first set of readings taken on day one. As with the CRT display, luminance output measurements over the three days were more variable than those taken on the same day.

Overall, the readings taken on the second day were higher in comparison to those taken on the first day, particularly at the left side of the display. The readings taken on the third day were very similar to those taken on the first day. The greatest luminance difference was noted at the upper left corner on both days, although the exact location differed slightly (a greater luminance output by 0.19 cd/m^2 and 0.03 cd/m^2 on days two and three respectively, a maximum variation of 1.98%). The minimum difference in luminance output was observed towards the right of the OLED display, at a location near the lower right corner on day two (a higher luminance output by 0.001 cd/m^2), and at a location near the upper right corner on day three (a lower luminance output by 0.00007 cd/m^2).
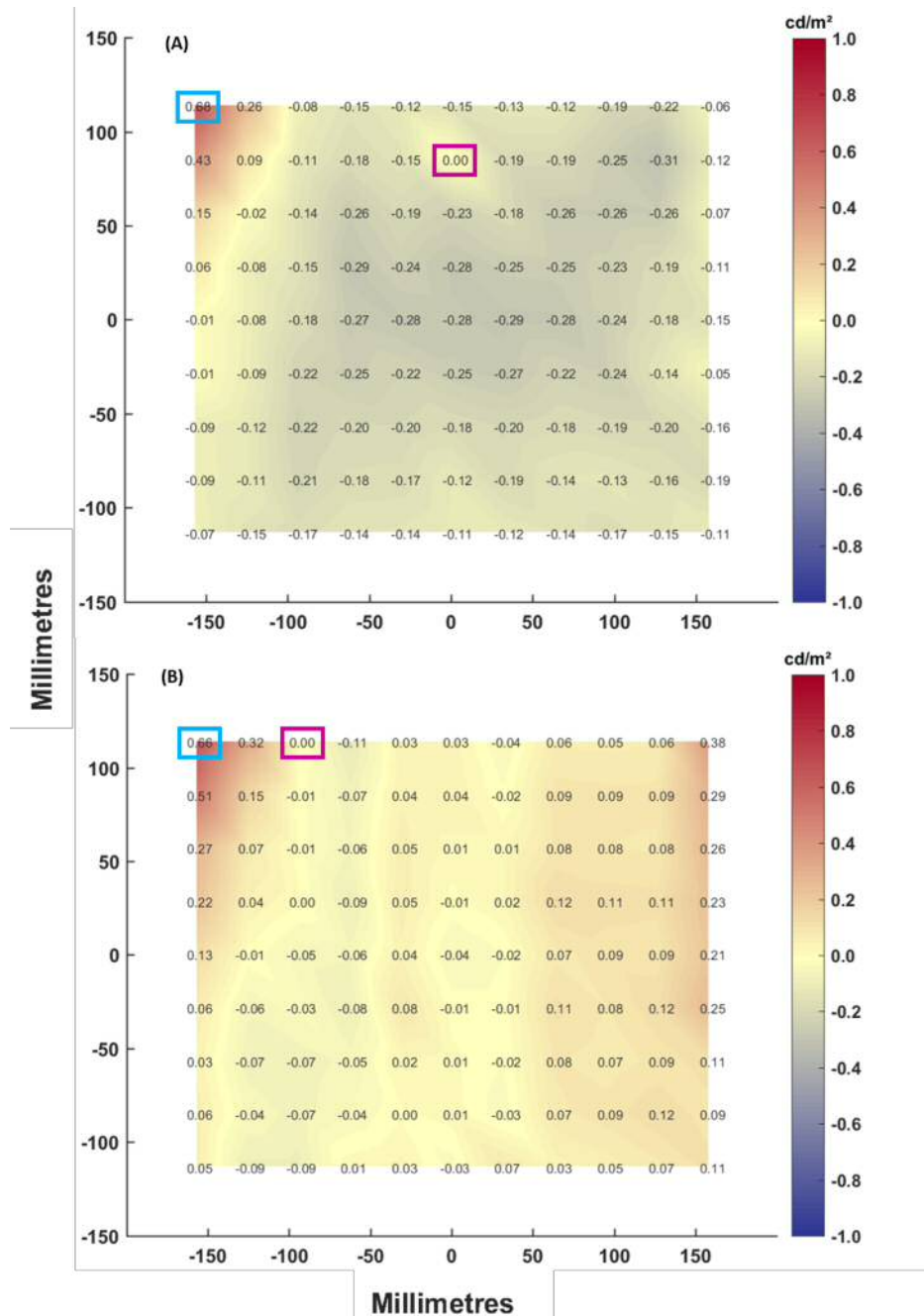
Figure 2.16 – Luminance difference in cd/m² on the (A) second and (B) third day, compared to the first set of readings on day one, with the Hewlett Packard p1230 chromatic CRT display. A positive value denotes a higher luminance output, and a negative value denotes a lower luminance output, when compared with the first set of readings. The x- and y- axes denote the co-ordinates for each location (mm). Luminance difference has been interpolated between the measured locations, and are indicated according to the colour scale on the right. The maximum luminance difference is indicated by the blue rectangle, and the minimum luminance difference is indicated by the magenta rectangle.
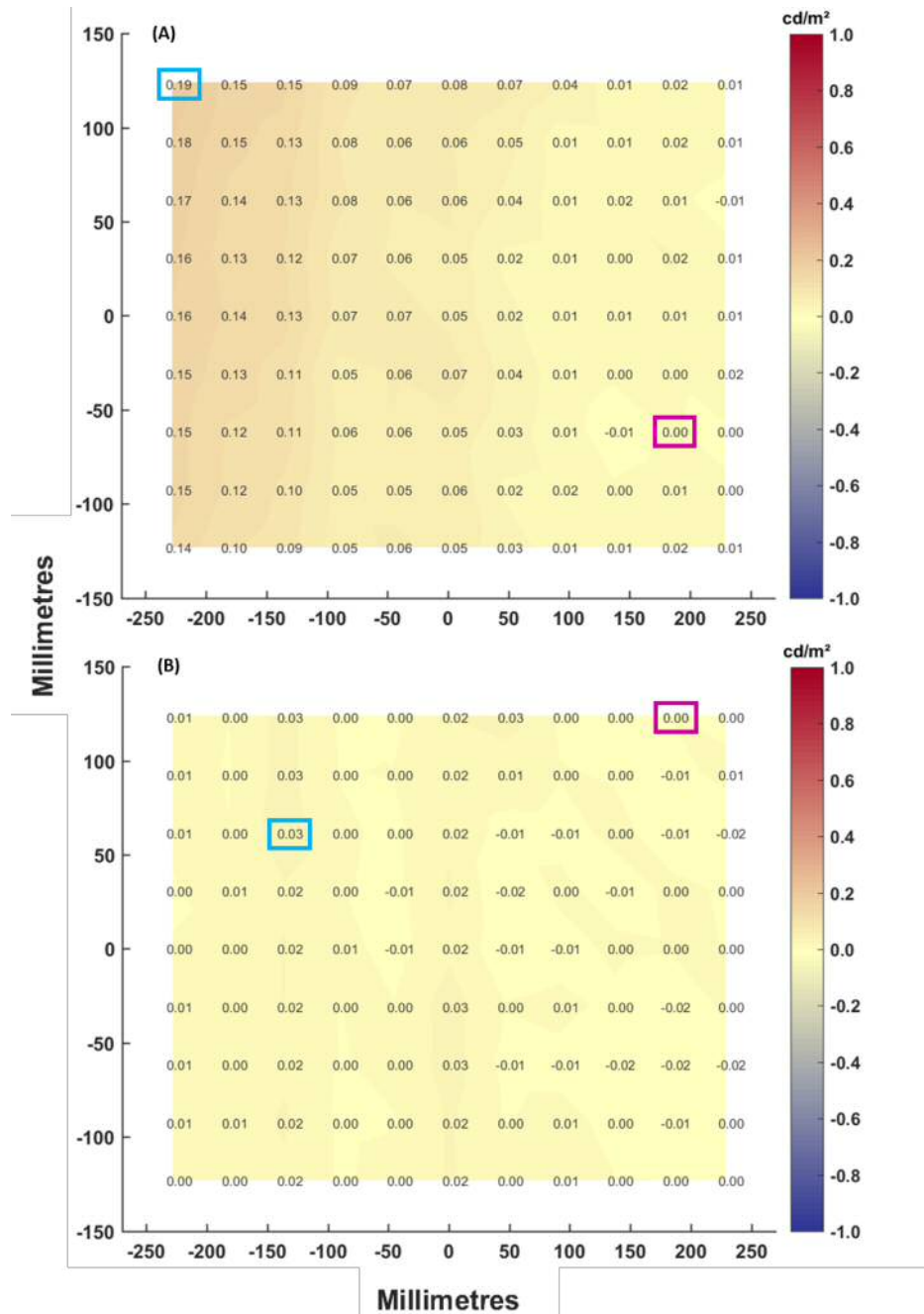
Figure 2.17 – Luminance difference in cd/m² on the (A) second and (B) third day, compared to the first set of readings on day one, with the Sony PVM-A250 Trimaster El OLED display. A positive value denotes a higher luminance output, and a negative value denotes a lower luminance output, when compared with the first set of readings. The x- and y- axes denote the co-ordinates for each location (mm). Luminance difference has been interpolated between the measured locations, and are indicated according to the colour scale on the right. The maximum luminance difference is indicated by the blue rectangle, and the minimum luminance difference is indicated by the magenta rectangle.

2.3.4 Experiment three - target dimensions

2.3.4.1 Method

To investigate whether target dimension differs across the display screen, thus impacting presented stimulus dimensions, circular outlines of 10 mm diameter (78.54 mm^2) were presented at different locations across the display (*Figure 2.18*). The horizontal and vertical diameters of each target were measured using a 7x Peak Mini Comparater, a loupe with an incorporated reticule scale of 0.1 mm divisions. This was conducted for both the Hewlett Packard p1230 CRT display, and the Sony PVM-A250 Trimaster El OLED display.
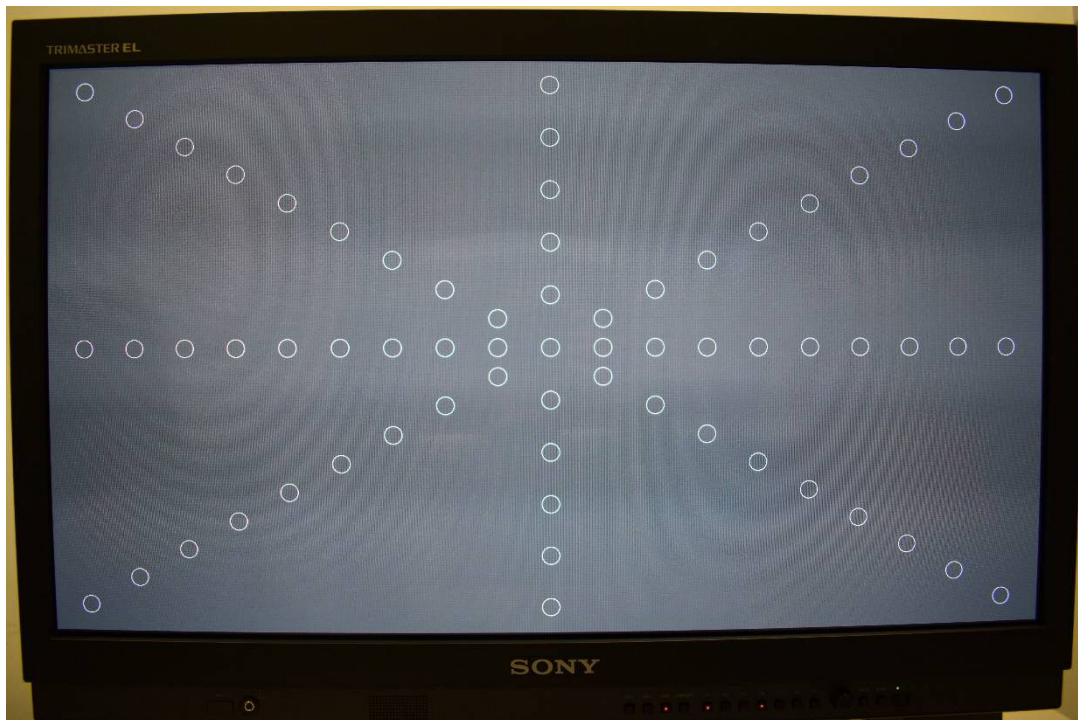
Figure 2.18 – Photograph showing the display on the Sony PVM-A250 Trimaster El OLED display, used to determine whether differences in stimulus dimension will occur across the display. This was also conducted with the Hewlett Packard p1230 chromatic CRT display.

2.3.4.2 Results

Results are displayed in *Figure 2.19* for the (A) CRT and (B) OLED displays. The x- and y- axes denote the co-ordinates of each location (mm), taken as the centre of each

target. Horizontal lines indicate the width and vertical lines indicate the height of each target. Line colour indicates the measured dimension, according to the scale on the right; this scale applies to both horizontal and vertical dimensions.
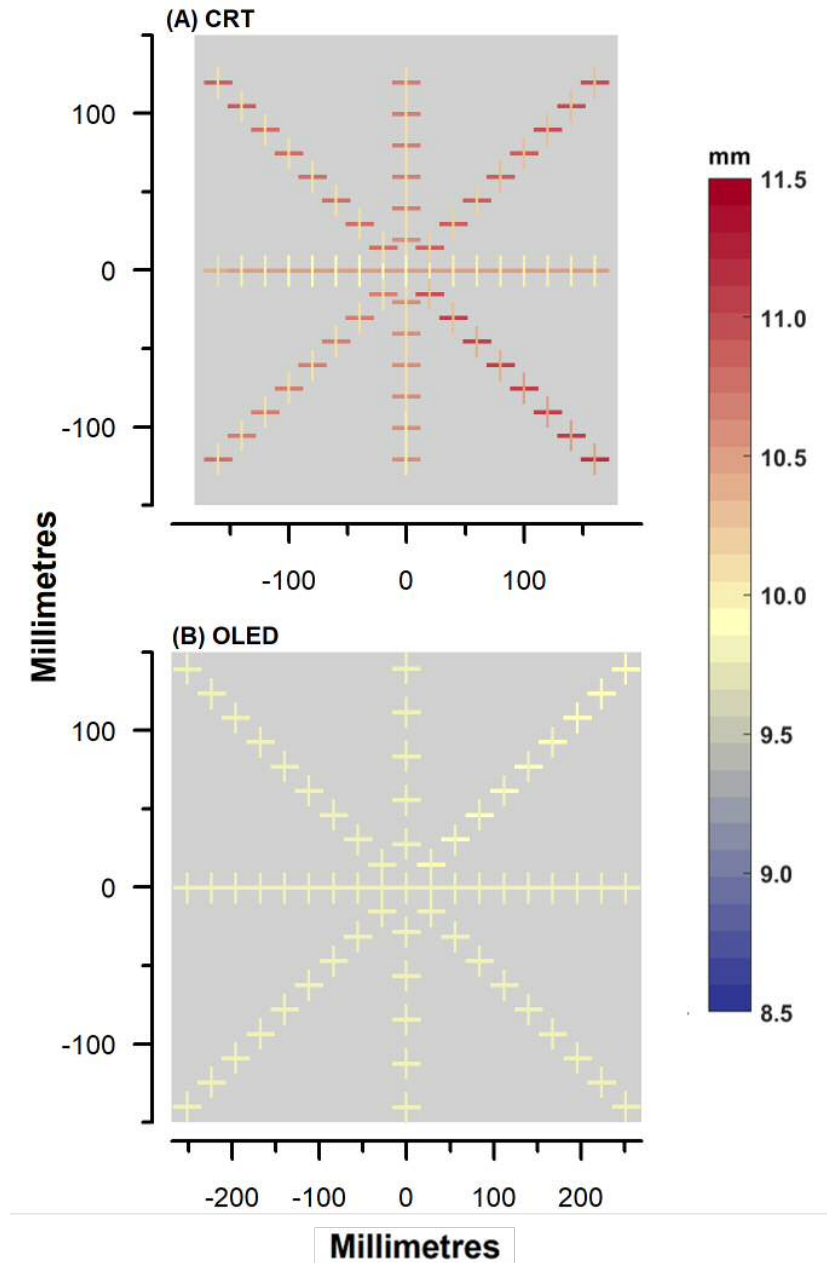
Figure 2.19 – Measured horizontal and vertical dimensions (mm). The x- and y- axes denote the co-ordinates for each location (mm). Horizontal lines indicate width and vertical lines indicate height of each target. Line colour indicates the measured dimension, according to the scale on the right.

With the CRT display (*Figure 2.19.A*), all horizontal dimensions were greater than that intended, by a maximum of 1.2 mm (12%), and a minimum of 0.4 mm (4%). Generally, horizontal dimensions were noted to increase with eccentricity from the centre of the CRT display, and were greater on the left side of the display compared with the right. Some vertical dimensions were greater than that expected, by a maximum of 0.5 mm (5%), and some were lower, by a maximum of 0.3 mm (3%). Vertical dimensions appeared less affected by distance from the centre of the display, although an increase was noted in the inferior of the screen in comparison with the centre. Overall, a greater area compared with that intended was found at all locations, with a maximum increase in area of 12.17 mm^2 (15.50%), and a minimum increase of 3.10 mm^2 (3.95%), from the intended 78.54 mm^2.

With the OLED display (*Figure 2.19.B*), all horizontal and vertical dimensions were lower than intended, by a maximum of 0.2 mm (2%), and a minimum of 0.1 mm (1%) in both dimensions. Neither horizontal nor vertical measurements appeared to be influenced by their eccentricity. Overall, a lower area than that intended was found at all locations, with a maximum decrease in area of 3.11 mm^2 (4.0%), and a minimum decrease of 2.34 mm^2 (3.0%), from the intended 78.54 mm^2.

2.3.5   Discussion

An ideal display screen would reach luminance stability quickly and maintain this stability throughout all tests, whether conducted on the same day or on different days. It would also maintain a uniform luminance output across the full display, referred to as spatial homogeneity, and maintain a constant target dimension, such that any displayed stimuli are of the same area, irrespective of their location on the display screen. An ideal display screen would also maintain a constant duration between frames, to ensure a constant stimulus duration, and luminance would remain independent of viewing angle.

2.3.5.1 Luminance stability

Time taken to achieve luminance stability with the Hewlett Packard p1230 CRT display is in keeping with times reported in the literature (Metha et al. 1993; Krantz 2000; Brainard et al. 2002; Klein et al. 2013), although once thermal equilibrium had been

achieved at 115 minutes, luminance stability remained within ±0.5% for the subsequent 4.08 hours, which is a slightly lower fluctuation than might be expected from the literature (Metha et al. 1993). It would therefore be beneficial to allow at least 115 minutes for luminance stabilisation, although luminance stability is at an acceptable level at 102 minutes.

Time taken to achieve luminance stability with the Sony PVM-A250 Trimaster El OLED display is again largely in keeping with the literature (Ito et al. 2013), although a slight difference in time taken to achieve equilibrium was noted on the two separate days. As fluctuations remained well within ±0.5% of the required luminance output after 28 minutes, and remained so for the subsequent 5.55 hours, it would be beneficial to allow at least 28 minutes for equilibrium to occur prior to conducting any experimental tests, although luminance stability is at an acceptable level after 16 minutes.

2.3.5.2 Spatial inhomogeneity

The variation in luminance output across the CRT display was generally in keeping with that of other studies, i.e. a reduction in luminance output was noted with increasing eccentricity from the centre of the display (Metha et al. 1993; Krantz 2000). The range of non-uniformity was not constant, with some days showing greater variation across the CRT display than others. However, the lower part of the screen did display a higher luminance output than the centre in the extreme periphery, which was perhaps unexpected. Given that the reduction in luminance output towards the peripheral edges has been attributed to the angle of the electron beam through the shadow mask (Metha et al. 1993), the higher luminance output at the lower peripheral edges may indicate a misalignment of one or more electron guns, or the shadow mask itself. A previous study, also noting a higher luminance output at the peripheral edge of their CRT display, hypothesised that this was due to the age of the monitor, or trauma-related disturbance in the electron guns (Mulholland 2014). The findings presented here are unlikely to be due to age, given that the CRT display was less than six months old at the time of the measurements. Additionally, there were no recorded incidences of trauma, therefore there is no reason to believe that the effect is due to trauma-related misalignment. This finding cannot be attributed to a particular CRT model, as the monitor used here is different to that used by Mulholland (2014).

The variation in luminance output across the OLED display was much less than that of the CRT display, in keeping with that of other studies (Ito et al. 2013), and the range of non-uniformity was less variable across different days. In addition, likely due to the pixel independence of the OLED display, variations in luminance output were largely unrelated to the eccentricity of the tested location.

It is worth noting that luminance output stabilised at a higher value than intended for the CRT display. The luminance output of any target presented on a display is achieved by inputting values between 0 and 1 for the red, green and blue sub-pixels (RGB value). The higher the value, the brighter that sub-pixel will be presented, and equal values for each of the sub-pixels will result in an achromatic presentation. The RGB values used to present the ~10 cd/m^2 background luminance on each display screen was achieved through preliminary testing prior to conducting the experiments presented here. Luminance stability was permitted prior to determining the relevant RGB values, in accordance with the warm-up times reported in existing literature for each display type. The CRT display was permitted 60 minutes, and the OLED display was permitted 20 minutes to achieve luminance stability, prior to conducting the preliminary tests to determine these RGB values. This may explain the higher-than-intended luminance output with the CRT display; as demonstrated in section *2.3.2*, this was not a sufficient time to complete thermal equilibrium, and luminance stability would not have been achieved within this time period. An adjustment of the RGB values, following sufficient time to permit luminance stability, would likely have achieved a luminance output closer to the intended 10 cd/m^2. However, as raw luminance output measures were not the focus of these experiments, but the variation between measurements, the findings presented here are still valid for both luminance stability and spatial inhomogeneity.

2.3.5.3 Target dimensions

It should be noted that, as measurements of target dimension were taken manually, some observer error is unavoidable. This is particularly true as images presented on either a CRT or an OLED display are not situated on the surface of the screen, and are therefore subject to parallax on observation, creating difficulties with measurement accuracy. Steps were taken to minimise these effects; measurements were taken by

one observer only and efforts were made to maintain consistency in measurement technique across both types of display.

Measured dimensions on the CRT display were dependent on eccentricity from the centre of the display, in which an increase in measurement was demonstrated with increasing distance from the centre of the display. This was true of both horizontal and vertical dimensions, although was more pronounced for horizontal dimensions. Given the structure of a CRT display, this is not unexpected, as electron beams will pass obliquely through the apertures in the shadow mask, creating a more pronounced, horizontal pixel bleed at the peripheral edges of the display screen.

Measured dimensions on the OLED display appeared largely independent of eccentricity, and were more consistent across the display screen than observed with the CRT display. Given the differences in structure and image generation between these two displays, this is not unexpected, as the control of pixels is largely independent in an OLED display, and they are not activated by a scanning electron beam through an aperture. The differences in stimulus dimension between the two types of display also confirm that the CRS toolbox does not apply a correction factor to the stimulus dimensions to account for the location from the centre of the screen (i.e. to account for viewing angle).

Ideally, two further experiments would also have been conducted, one to investigate luminance difference with viewing angle, and one to investigate phosphor activation. Studies have indicated that luminance is affected by viewing angle with an OLED display (Ito et al. 2013), but that luminance is largely independent of viewing angle with a CRT display (Ghodrati et al. 2015). It would therefore have been beneficial to measure this with the CRT and OLED displays utilised here, however this was not possible due to limitations in available instrumentation, as the ColorCAL II is unsuitable for this type of measurement. However, if the study of Ito et al. (2013) is considered, a luminance reduction is not expected within 10° of fixation, and a reduction of ~13% is expected at 30° from fixation, the maximum eccentricity tested in the experiments presented in this thesis. If both direct luminance measurements, as investigated here, and expected luminance reduction with viewing angle are considered, a lower spatial

inhomogeneity would still be observed with the OLED display than with the CRT display.

The examination of phosphor activation would permit an analysis of temporal precision, which affects stimulus duration. Stimulus duration has been shown to be largely dependent on the methods by which it is calculated, and is more accurate when phosphor decay time is considered (Bridgeman 1998; Mulholland et al. 2015d). However, in this series of experiments, the exact stimulus duration is not crucial, but must be consistent across all presented stimuli. Studies show a consistent luminance profile with both CRT and OLED phosphors, independent of luminance, which will result in a consistent stimulus duration (Bridgeman 1998; Cooper et al. 2013; Elze et al. 2013; Ghodrati et al. 2015; Mulholland et al. 2015d). Given the higher duty cycle of the OLED phosphors (Cooper et al. 2013; Elze et al. 2013), the presented stimulus duration may be more similar to the intended stimulus duration in an OLED display than with a CRT display, although adjustments could be made to account for this in the experimental input parameters.

Overall, the results presented here for all experiments confirm that the OLED display is the more appropriate display to use in the series of experiments presented in this thesis. The OLED display demonstrates many advantages over the traditionally used CRT display, in the form of a shorter time to luminance stabilisation, reduction of spatial luminance inhomogeneity, reductions in short- and long-term fluctuations of luminance output, and greater consistency of target dimensions across the display. Although a reduction in luminance with viewing angle will occur, studies indicate that this effect is less than the difference in luminance output noted with direct measurements of the CRT display. Although the refresh rate of the Sony PVM-A250 Trimaster El OLED display is fixed at 60 Hz, the flicker-free option permits operation at 120 Hz, in keeping with the recommendations for vision-based research (Shady et al. 2004; Zele and Vingrys 2005). As no presented stimulus will be > 40% of the screen, the restrictions in luminance output under these circumstances will not impact these experiments (Elze et al. 2013; Ito et al. 2013).

As such, any differences in thresholds between the stimulus paradigms investigated in this series of experiments, whether conducted on the same day or on different days, cannot be attributed to variations in the apparatus used, and may more confidently be attributed to the stimulus form itself.

Chapter 3 The psychometric function, and investigation of response variability characteristics with Goldmann I-V

This chapter discusses some of the considerations of the study design and fitting of statistical models, and describes a preliminary experiment that was conducted to investigate the response variability of Goldmann I-V stimuli in healthy observers.

3.1   The method of constant stimuli and psychometric function fitting

3.1.1   Introduction

Several experiments presented in this thesis utilised a MOCS procedure. This is a non-adaptive psychophysical technique, in which several stimulus values are pre-selected, ranging in visibility from never seen (i.e. 0% seen), to perceived every time (i.e. 100% seen). These stimulus values are presented to the observer many times in a randomised order. For each stimulus value, the number of presentations correctly responded to are recorded and used to construct a FOS curve. An example of this is shown in *Figure 3.1*, in which the characteristic sigmoidal shape is demonstrated, with asymptotes at markedly sub- and supra-threshold stimulus values (Corliss and Norton 2002).
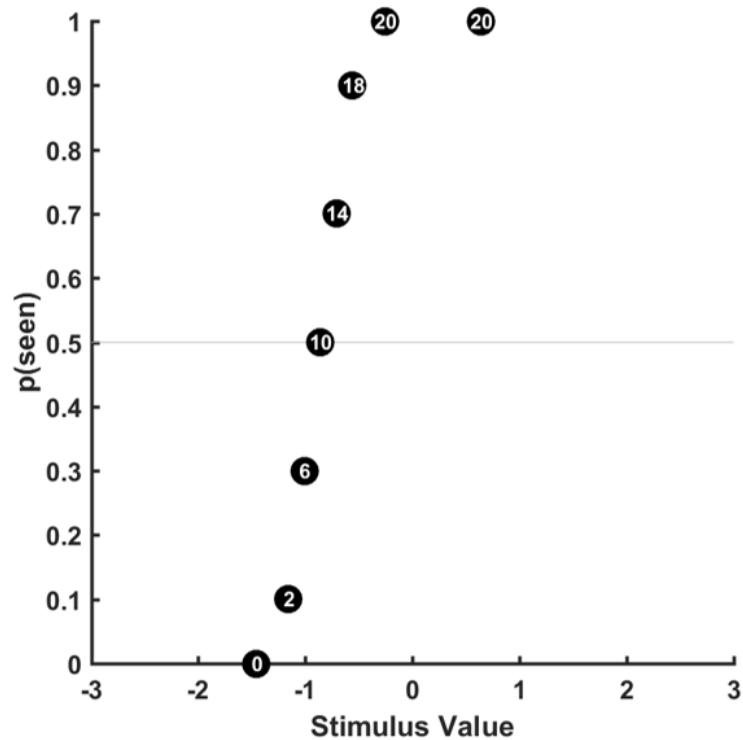
Figure 3.1 – Example FOS curve. Each stimulus level was presented 20 times. The value in each data point denotes how many of the 20 repetitions were responded to by the observer, and p(seen) indicates the proportion seen.

3.1.2   The psychometric function

The FOS data are typically fitted with a sigmoid-type function (psychometric function), a smooth curve which models the relationship between the physical stimulus and the observer's responses, and from which various characteristics may be estimated about the underlying sensory mechanism.

While an in-depth discussion of the psychometric function is beyond the scope of this thesis, it will be considered briefly here. The psychometric function is a means of fitting binary responses with a linear regression model, often by the method of maximum likelihood. This method has been refined over the years, and now describes the process by which the response data are linearised by a so-called 'link function', then fitted with a linear regression model. Various types of link function are available that make assumptions about the distribution of the data error (Treutwein and Strasburger 1999).

The general formula of a psychometric function $\psi(x)$ is shown in *Equation 3.1*.

$$\psi(x; \alpha, \beta, \gamma, \lambda) = \gamma + (1 - \gamma - \lambda)F(x; \alpha, \beta)$$

$\alpha = threshold$

$\beta = slope$

$\gamma = guess\ rate$

$\lambda = lapse\ rate$

Threshold (α) is generally determined as the point halfway between the upper and lower asymptotes, the position of which may differ depending on experimental design. For the purposes of the experiments presented in this thesis, consisting of a detection task in which the observer was instructed to respond to any stimulus they had detected in their visual field by pressing a button on a response pad, the threshold was taken as the point on the psychometric function where the stimulus was perceived 50% of the time, i.e. p(seen) = 0.5. Examples of psychometric functions with differing thresholds (indicated by the grey line), but the same slope, are shown in *Figure 3.2*.
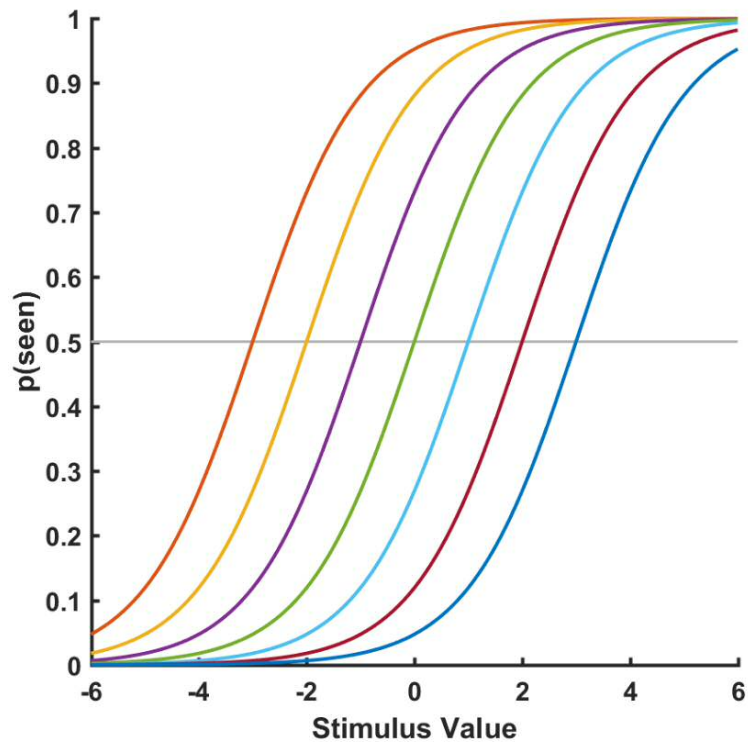
Figure 3.2 – Examples of psychometric functions, each with a different threshold (i.e. p(seen, 0.5), but with the same steepness of slope (MATLAB code courtesy of https://davehunter.wp.st-andrews.ac.uk/2015/04/12/ fitting-a-psychometric-function, accessed on 07/11/17).

The slope (β) of the function is indicative of the change in observer response with stimulus change. Examples of psychometric functions with the same threshold, but different slopes, are shown in *Figure 3.3*; a higher response variability is indicated by the shallower slopes (e.g. blue), and a lower response variability is indicated by the steeper slopes (e.g. orange). The expression describing the slope parameter differs between different types of psychometric function, which can make it difficult to compare between models, and consequently to compare results between studies. Strasburger (2001) and Gilchrist et al. (2005) advise three alternative methods of expressing the slope value, such that direct comparisons may be made between psychometric function models, two of which are commonly found in the literature for vision-based research. One method is to determine the local gradient at the location of threshold (Wichmann and Hill 2001a; Kingdom and Prins 2009), and the other method is to express the response variability as the spread of the data, i.e. the stimulus range between two, nonasymptotic points on the psychometric function

slope; this may be taken as the interquartile range (IQR, Chauhan et al. 1993; Strasburger 2001), or the standard deviation (SD, McKee et al. 1985; Treutwein and Strasburger 1999; Henson et al. 2000; Kingdom and Prins 2009; Turpin et al. 2010).
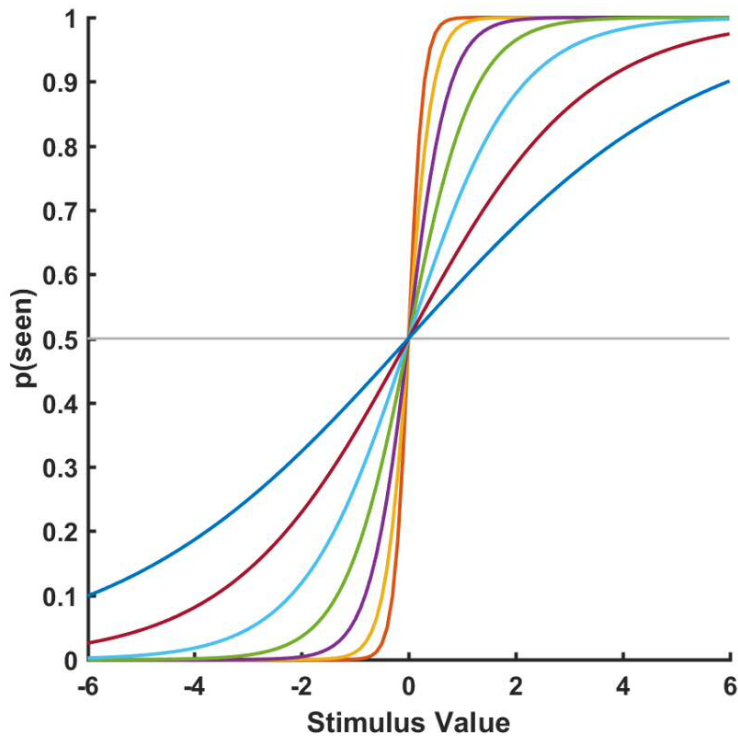
Figure 3.3 – Examples of psychometric functions, each with the same threshold (i.e. p(seen, 0.5), but differing steepness of slope (MATLAB code courtesy of https://davehunter.wp.st-andrews.ac.uk/2015/ 04/12/fitting-a-psychometric-function, accessed on 07/11/17).

The guess rate (γ) and lapse rate (λ) describe chance-level performance, and attentional lapse respectively. While threshold and slope estimate properties of the underlying sensory mechanism, guess and lapse rates describe how a non-perfect observer may respond, and are necessary to fully specify the psychometric function (Kingdom and Prins 2009). In a detection MOCS experiment, in which an observer responds to a stimulus they have detected, but no action is taken if no stimulus is perceived (as used in the experiments described in this thesis), the guess rate describes an occurrence whereby the observer responds to a stimulus that is sufficiently sub-threshold to be beyond the detection of the visual system. This would

be described as a 'false positive' in a clinical setting, usually attributed to internal (within the visual system) or external (within the apparatus) noise. The lapse rate generally describes those occurrences whereby an observer fails to respond to a stimulus that is sufficiently supra-threshold as to be well within the detection capabilities of the visual system, such that a response is always expected. This would be described as a 'false negative' in a clinical setting. These parameters are often specified as zero prior to fitting the psychometric function model, giving no allowances for guessing or lapsing, however studies have shown that an inappropriate selection of the guess/lapse rates can cause a substantial change in the fit of the model, leading to considerable biases in the estimates of threshold and response variability (Treutwein and Strasburger 1999; Wichmann and Hill 2001a). Therefore, although a researcher may not wish to analyse the guess and lapse rates, they should be considered in the fitting of a psychometric function, as it may be inappropriate to assume a perfect observer.

This is of particular importance if an incomplete FOS curve is expected. In an experiment such as that described in chapter four, in which some participants had a reduction in sensitivity due to the presence of glaucoma, the term 'lapse rate' is something of a misnomer. It may not be possible to measure a complete FOS curve, as there may not be an available stimulus that is sufficiently supra-threshold for the observer to perceive every time it is presented. This is observed in clinical practice, where a high false negative rate may be indicative of disease, rather than inattention (Bengtsson and Heijl 2000). As such, in the experiments undertaken here, the fitting of the psychometric function model must be robust to an incomplete FOS curve.

3.1.3 Minimising bias of the psychometric function

Inappropriate stimulus selection may result in an unnecessarily incomplete FOS curve, due to insufficient sampling of the sub- or supra-threshold range of the curve. In addition, the number of trials, and their position on the FOS curve, has been shown to influence the resulting threshold and slope/spread parameters established from the psychometric function (Hill 2001). Care should be taken when comparing between stimulus forms to ensure that equal numbers of sub- and supra-threshold values are presented with each stimulus form, as greater visibility of one stimulus over another,

due to a greater number of supra-threshold values, can artefactually steepen the psychometric function.

Hill (2001) investigated seven sampling schemes (the position of the stimulus values on the FOS curve) for a yes-no, and seven for a two alternative forced choice (2AFC) psychophysical methodology. Although not identical to the detection task utilised in the MOCS experiments presented in this thesis, the yes-no design is a close approximation, and the findings of Hill (2001) were used to inform the stimulus selection utilised in these experiments. An ideal sampling scheme would result in a precise estimate of threshold and slope, with minimal bias. The work of Hill (2001) demonstrates that, while one sampling scheme may be optimal for threshold estimation, it may be sub-optimal for slope estimation, and vice versa. Hill (2001) additionally investigated the number of stimulus values, and the number of repetitions of each value, to determine what effect this had on threshold and slope estimate. As this investigation was done via computer-simulated models, one must also consider the effect of fatigue, as test duration will affect the accuracy of observer responses.

Considering all these factors together, the sampling scheme chosen for the experiments presented in this thesis comprised of eight stimulus values, located along the FOS curve according to *Figure 3.1* (plotted for an ideal observer), and each repeated 20 times (160 presentations). This sampling scheme represented a compromise between the ideal factors, such that both threshold and slope/spread of the psychometric function could be estimated with satisfactory accuracy.

As the presented stimulus values are selected prior to conducting the experiment, some prior knowledge is required of the expected FOS curve to ensure accurate placement of the stimulus values. To achieve this, a two-stage protocol was developed in which an initial short MOCS procedure was conducted prior to a standard MOCS procedure. The short MOCS procedure enabled estimation of the position and shape of the FOS curve, and was used to inform stimulus selection for the standard MOCS procedure. The experiment in section *3.2* describes this two-stage protocol, and investigates its utility.

3.2 Two experiments analysing the response variability of Goldmann I-V in healthy observers

3.2.1 Introduction

Previous studies have reported that the use of a smaller perimetric stimulus area results in greater variability of the sensitivity, while the use of a larger perimetric stimulus area results in lower variability of the sensitivity (Wall et al. 1997; Wall et al. 2013). As the series of experiments presented in this thesis involves the use of small stimuli, and stimuli of non-constant area, this is an important factor to consider.

As discussed in section *3.1*, an examination of the work of Hill (2001) reveals that, while MOCS can be a very robust technique, the predetermined stimulus values must be selected carefully to prevent unduly influencing the resulting threshold/sensitivity and response variability. An examination of Wall et al. (1997), in which MOCS was used to determine variability characteristics of Goldmann I, Goldmann III and Goldmann V, highlights some aspects of their methodology which could have unduly influenced the results, such that the lower variability found with the Goldmann V stimulus may not be due to the stimulus itself.

To determine the dB values presented for the Goldmann III stimulus, Wall et al. (1997) used the perimetric sensitivity established using the Humphrey Field Analyzer (either the 24-2 or 30-2 Full Threshold strategy), plus 10 dB higher and lower, at 2 dB intervals. To determine the dB values presented for the Goldmann I stimulus, 10 dB was subtracted from the stimulus values used for the Goldmann III stimulus. For the Goldmann V stimulus, values ranged from 20 to 42 dB, in intervals of 2 dB.

There are several factors of note with this stimulus selection. A comparison of stimuli with a non-clinical technique, such as MOCS, is difficult if one is restricted to using integer dB steps with clinical instruments. Sampling the psychometric functions for differing stimulus areas in 2 dB steps will lead to supra-threshold stimuli being more visible with the Goldmann V stimulus than with smaller stimuli, due to the greater raw energy steps between nominal contrast levels. This greater visibility may well lead to an artefactual steepening of the psychometric function, and thus an apparently lower response variability with a larger stimulus area.

Establishing stimulus values for a Goldmann I stimulus by subtracting 10 dB from the stimulus values used with a Goldmann III stimulus, irrespective of eccentricity, is somewhat arbitrary. As discussed in section *1.4*, area and contrast are inversely proportional at threshold within the area of complete spatial summation, however beyond Ricco's area this relationship no longer holds. Given the known increase in Ricco's area with increasing eccentricity from fixation (Graham and Bartlett 1939; Graham et al. 1939; Hallett 1963; Wilson 1970; Vassilev et al. 2003), and the known differences in Ricco's area between glaucoma and healthy observers (Sloan 1961; Dannheim and Drance 1974; Redmond et al. 2010a), in addition to the individual variability in Ricco's area values, it is likely that this approach serves as an appropriate estimate of perimetric sensitivity for some observers, at some visual field locations (if both Goldmann I and Goldmann III are within Ricco's area), but does not serve as an appropriate estimate of perimetric sensitivity for other observers/visual field locations (if Goldmann I is within Ricco's area, but not Goldmann III). For the Goldmann III stimulus, although the stimulus values used in the MOCS procedure are based on the sensitivity measure from HFA, it is likely that the visual range from p(seen) = 0, to p(seen) = 1, differs in healthy observers compared to those with glaucoma. Using the same range for both cohorts may result in incomplete FOS curves in some observers compared with others. For the Goldmann V stimulus, the same dB range was used for all observers, without considering the individual's visual sensitivity. Again, this is likely to result in incomplete FOS curves in some observers, or may present too many stimulus values at the p(seen) = 1 level for other observers, leading to sub-optimal sampling of the rest of the FOS curve.

It is therefore likely that differing numbers of supra- and sub-threshold stimulus values were presented with different stimulus areas, and that these also differed between the healthy and glaucoma cohorts. As demonstrated by Hill (2001), the position of the stimulus values along the FOS curve can influence the threshold and response variability determined from the psychometric function. If differing numbers of sub- and supra-threshold stimulus values were presented with different stimuli/cohorts, this could unduly influence the threshold and response variability with one stimulus/cohort over another. Indeed, an observation of the example psychometric

functions provided, Figure 2 in Wall et al. (1997), does suggest an uneven sampling of the FOS curve between different stimulus areas, and between cohorts.

It is also worth noting that different Goldmann stimuli will sample different sections of the dB scale. As noted in section *1.3.2*, a 30 dB Goldmann III stimulus contains less raw energy than a 30 dB Goldmann V stimulus. Additionally, a 2 dB interval with a Goldmann III stimulus, represents a smaller change in raw energy than with a Goldmann V, which will operate higher on the dB scale.

As such, it could be that the differing response variabilities found between Goldmann I, III and V in Wall et al. (1997) were due to sub-optimal stimulus selections for the MOCS experiment, and not due to the stimulus configuration itself.

This was investigated in the two experiments presented here, in which a two-phase protocol was developed, informed by the studies of Hill (2001), to aid in the pre-selection of stimulus values for a MOCS procedure. Presented stimuli were not limited to integer dB values, to permit finer stimulus increments, and to reduce some of the differences in raw energy step size between stimuli due to their differing positions on the dB scale. Tests were conducted with all five Goldmann stimuli to further investigate whether response variability was dependent on stimulus area.

3.2.2 Methods, experiment one – investigation of sensitivity and response variability at differing eccentricities

3.2.2.1 Participants

Five young, healthy participants were recruited to this experiment, aged 24.6, 25.1, 26.1, 26.9, 30.4, and 48.0 years. One eye of each participant was tested. This was selected as the eye that best met the inclusion/exclusion criteria, or was selected as the right eye if both eyes were equally suitable. For all five participants, the right eye was the test eye. All participants underwent testing with the SITA Standard 24-2 program on the Humphrey Field Analyzer (HFA II, Carl Zeiss Meditec Inc., Dublin, CA); this was performed twice with the test eye to ensure adequate perimetric experience prior to undertaking any experimental tests. All participants had a normal visual field ('within normal limits' according to the Glaucoma Hemifield Test), with a median [IQR]

MD of -1.2 dB [-1.4, -0.8]. False positive and false negative rates were < 10% for all participants.

Participants did not have any systemic/ocular disease and/or medications known to affect visual performance (e.g. diabetes, hydroxychloroquine medication). Ocular health was confirmed via slit lamp biomicroscopy, and no participant had previously undergone any ocular surgery.

All participants had a best-corrected visual acuity of 6/6 or better in the test eye, in the absence of any corneal or media opacities. Refractive error was restricted to < ± 3.00 DS, and < -1.50 DC. This refractive error range is more limited than is conventionally imposed in perimetric research, due to the eccentricity of some of the test locations; the use of corrective lenses was not feasible, as some locations were beyond the standard 30° (*Figure 3.4*). By restricting refractive corrections as described, this ensured an adequate, unaided visual standard. Of the five participants tested, one participant had a slightly hyperopic spherical correction (+0.25 DS), one was emmetropic, and three were myopic (≤ -1.00 DS) in the test eye. Astigmatism was ≤ 1.00 DC for all participants. One participant was presbyopic, with a distance refractive error of plano/-0.75 x 180; although this participant occasionally used a correction of +1.50 DS for near vision, they were still able to accommodate by 2.5 D. All tests were undertaken with natural pupils.

Ethical approval for the study was obtained from the School of Optometry and Vision Sciences Research and Audit Ethics Committee, Cardiff University. The research adhered to the tenets of the Declaration of Helsinki. Written, informed consent was obtained from all participants prior to inclusion.
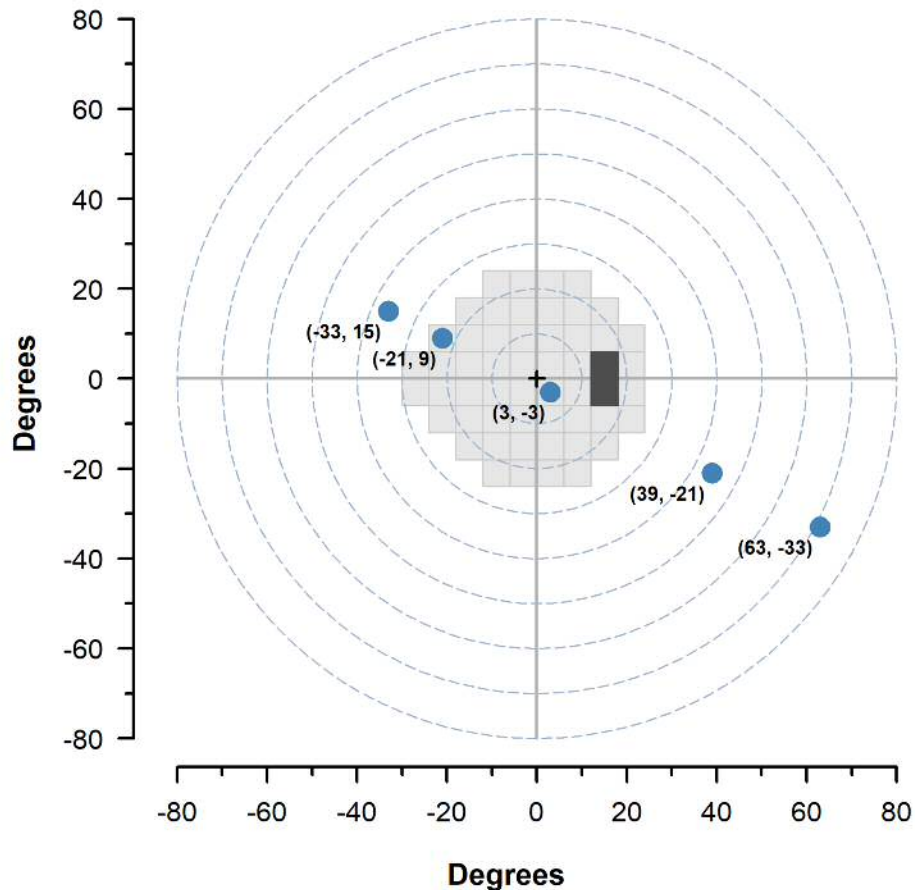
Figure 3.4 – Test locations for a right eye, corresponding to (L-R) 36.2°, 22.8°, 4.2°, 44.3°, and 71.1° from fixation. A standard 24-2 test grid is shown in grey for reference.

3.2.2.2 Apparatus and stimuli

As this experiment used only Goldmann stimuli, and test locations extended to 71.1° from fixation, the most suitable instrument was a commercially-available projection perimeter. This was the only experiment presented in this thesis conducted on a commercial perimeter. Stimuli were presented on an Octopus 900 perimeter (Haag-Streit, Koeniz, Switzerland), driven by the freely available, open source statistical environment R (R Development Core Team, 2014) and associated package 'OPI' (Turpin et al. 2012; Turpin 2013). R code was adapted for this experiment from that of T. Redmond. A nominally uniform background luminance of 10 cd/m^2 was maintained, in keeping with that conventionally used in SAP.

Stimuli were presented at five test locations in the supero-nasal, and infero-temporal quadrants, ranging from 4.2° to 71.1° from fixation (*Figure 3.4*), to incorporate a wide

range of sensitivity measures. Although test locations were not equidistant, they were positioned such that attentional bias should be minimal.

Each participant undertook tests for each of the five Goldmann stimuli. Goldmann I-V are achromatic, circular, incremental stimuli, varying in contrast. Areas for these stimuli are displayed in *Table 3.1*. Stimuli were presented with a fixed duration of 0.2 seconds.

Goldmann stimulus	I	II	III	IV	V
Area (log deg^2)	-2.02	-1.42	-0.84	-0.24	0.37

Table 3.1 – Goldmann stimulus I-V, with corresponding areas in log deg^2.

3.2.2.3 Psychophysical procedure

FOS curves were determined at each of the five test locations using a MOCS procedure, conducted separately for each of the five Goldmann stimuli. As discussed in section *3.1*, a MOCS procedure involves the presentation of a small number of pre-selected stimulus values, each one displayed multiple times throughout the test. A FOS curve is created by establishing the proportion of each stimulus value responded to by the observer. As discussed, the choice of presented stimulus values may unduly influence the shape of the curve if not selected appropriately. In an attempt to maximise efficiency, minimise bias of the response variability, and equate stimulus visibility across all conditions, a protocol was adopted whereby the psychometric function was densely sampled around the expected p(seen) = 0.5 region, guided by the work of Hill (2001), with sufficiently supra- and sub-threshold stimuli presented on the expected asymptotes. To achieve this, some prior knowledge of the individual observer's FOS curve at each of the tested locations was required, therefore a two-phase protocol was developed to aid in stimulus selection, as described below.

Phase one - short MOCS phase

The purpose of this phase was to establish an approximate position and shape of the FOS curve, in order to optimise sampling of the curve in the next phase (standard

MOCS phase). This phase consisted of a short MOCS procedure using nine stimulus levels, each presented five times, at each of the five test locations (225 presentations in total).

An estimated perimetric sensitivity had previously been established at each of the five test locations for each of Goldmann I-V with the HFA II; these data were from a series of pilot studies, conducted prior to the commencement of this PhD (Redmond and Artes 2012). The nine stimulus levels consisted of this perimetric sensitivity (the expected p(seen, 0.5) value), plus three above and three below this level in 1.5 dB intervals. An additional two levels were selected 9 dB above and below the initial stimulus level. Upper stimulus levels were capped at 40 dB, i.e. where the estimated p(seen, 0.5) level was > 31 dB.

Presentations were randomised in terms of stimulus level and test location. Participants were instructed to fixate the central target, and respond to any stimulus they had detected in their visual field by pressing the response button. Participant fixation was monitored throughout the test via the video display of the Octopus 900.

This short MOCS phase was of approximately 10 minutes duration, with a rest break taken at halfway (after 112 presentations). A FOS curve was constructed from the observer responses for each of the five test locations, and fitted with a cumulative Gaussian psychometric function.

Stimulus levels at p(seen) = 0.1, 0.3, 0.5, 0.7, and 0.9 were estimated from the psychometric function, and used to establish the stimulus levels presented in phase two, a theoretical example of which is shown in *Figure 3.5.A*.
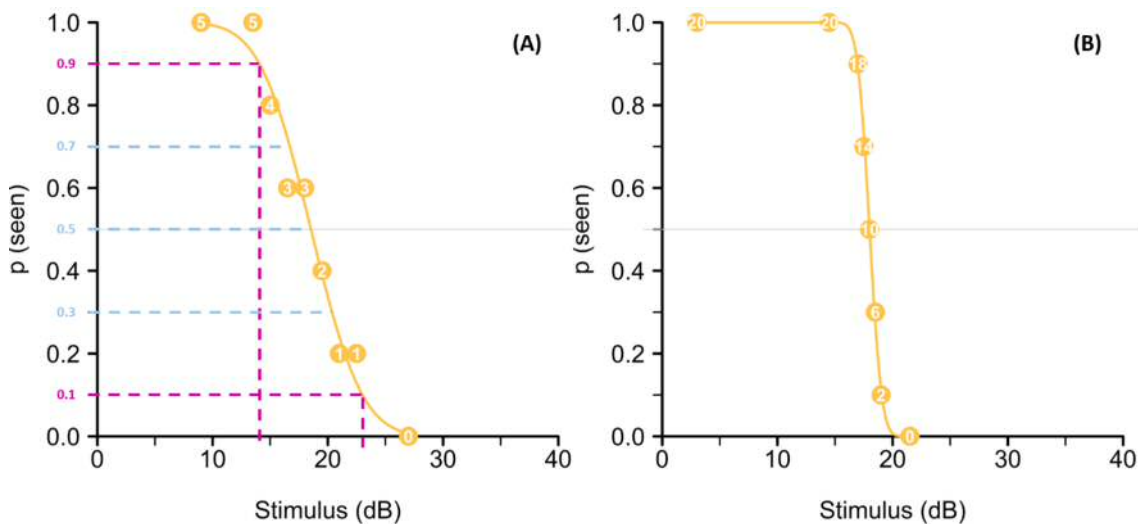
Figure 3.5 – Theoretical examples of FOS curves for (A) phase one, short MOCS phase, and (B) phase two, standard MOCS phase.

Phase two - standard MOCS phase

Five repetitions are generally regarded as too few for a MOCS experiment, however the purpose of the short MOCS phase was to approximate the FOS curve only. While the same, previously established stimulus levels were presented to all participants in the short MOCS phase, this was not so for the standard MOCS phase; stimulus visibility was equated across the five Goldmann stimuli, and across the five test locations for all participants.

In the standard MOCS phase, participants were presented with 20 repetitions of eight stimulus levels at each of the five test locations (800 presentations in total). At each location, five stimulus levels were determined from the FOS curve constructed in the short MOCS phase, the values for p(seen) = 0.1, 0.3, 0.5, 0.7, and 0.9. The interval between these stimulus levels was termed the 'step interval', and the minimum stimulus interval was limited at 0.5 dB. An additional two stimulus levels were established as p(seen, 0.5) ± step-interval*5, or p(seen, 0.5) ± step-interval*7 if the step interval was 0.5, ensuring an adequate range of stimulus levels. An extra, brighter stimulus level was established as p(seen, 0.5) -15 dB, to ensure a greater number of supra- than sub-threshold stimulus levels, thus aiding observer attention. As in phase

one, presentations were randomised in terms of stimulus level and test location. Upper stimulus levels were capped at 40 dB.

This standard MOCS phase was of approximately 40 minutes duration, with a rest break at every quarter (after 200 presentations). In both the short and the standard MOCS phases, participants could request additional rest breaks as required.

A FOS curve was again constructed from the observer responses for each of the five test locations, and fitted with a cumulative Gaussian psychometric function, a theoretical example of which is shown in *Figure 3.5.B*. Guess and lapse rates were permitted to vary.

Participants completed tests for each of the five Goldmann stimuli on separate days, with both phases for the same Goldmann stimulus completed on the same day. The order of the stimuli tested was randomised between participants, and all experimental tests were completed within a two-month period.

3.2.2.4 Statistical analysis

Fitting of psychometric functions, and analysis of FOS data were performed with the open source statistical environment R (R Development Core Team, 2014), using the 'psyphy' package (Knoblauch 2014). Other statistical analyses were performed with R or SPSS (IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0, Armonk, NY: IBM Corp). Analyses described from this point were conducted on those FOS data collected in phase two.

As this was a detection task, threshold was determined as p(seen) = 0.5. Response variability was determined as the spread of the psychometric function, taken as the SD. As 75th and 25th percentiles are located at mean $\pm$ 0.674 in a normative distribution, SD can be conveniently calculated from a psychometric function according to *Equation 3.2*.

$$SD = \frac{IQR}{1.349}$$

<div align="right">

Equation 3.2
(Silverman 1986)
</div>

$SD = Standard\ Deviation$

$IQR = Interquartile\ Range$

To examine the association with eccentricity, mean sensitivity for the five participants, i.e. p(seen) = 0.5, was plotted against eccentricity from fixation, ± one SD. This was conducted for each of the five Goldmann stimuli.

As for sensitivity, mean response variability for the five participants, i.e. SD of the psychometric function, was plotted against eccentricity from fixation, ± one SD, to examine the association between eccentricity and response variability. This was conducted for each of the five Goldmann stimuli to permit comparisons between differing stimulus areas. As response variability values were not normally distributed with all Goldmann stimuli at each test location (Shapiro-Wilk test), a Friedman analysis, with post hoc Wilcoxon signed-rank tests, was conducted to compare response variability values between the five Goldmann stimuli; this was performed individually at each of the test locations.

In all statistical analyses, a Holm-Bonferroni post hoc correction was applied where there were multiple tests of the same hypothesis. All p-values quoted here have been post hoc corrected were appropriate.

3.2.3   Methods, experiment two – investigation of sensitivity and response variability at equidistant locations in all four quadrants

One participant (age 48.0 years) completed a subsequent experiment, in which sensitivity and response variability with each of the five Goldmann stimuli were compared at equidistant locations in the four quadrants. Eight locations were tested, at 36.2° and 22.8° from fixation, in all four quadrants (*Figure 3.6*).
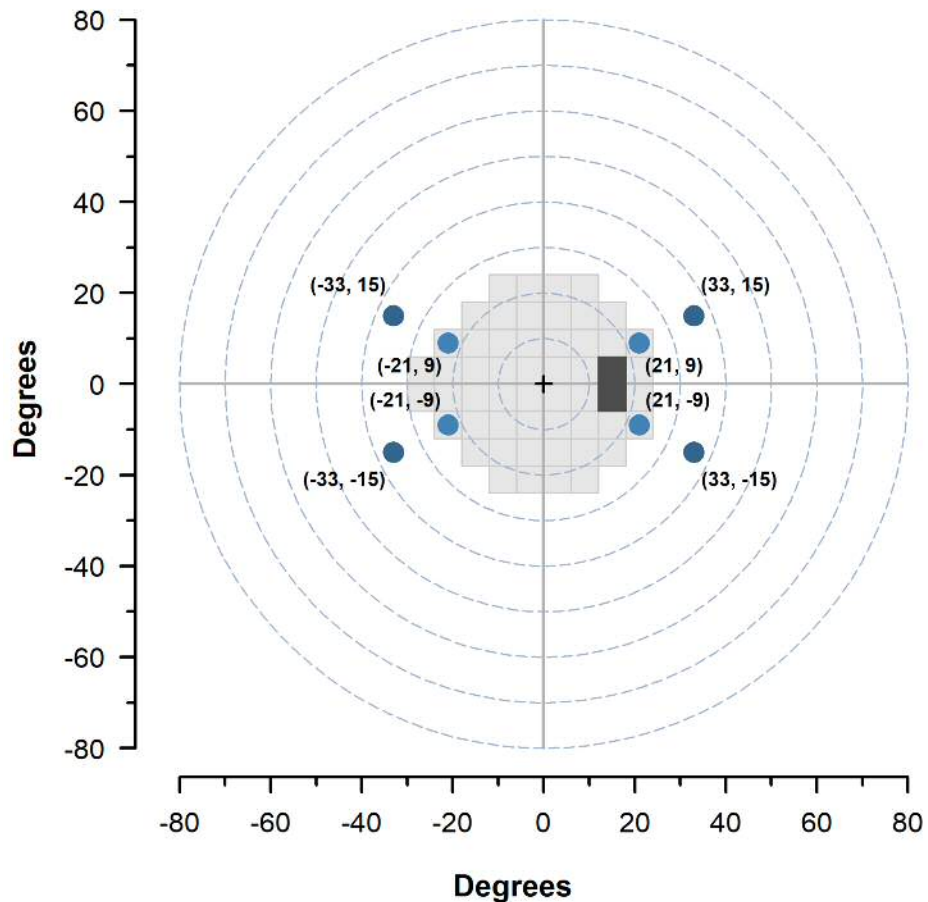
Figure 3.6 – Test locations for a right eye, corresponding to 36.2° and 22.8° from fixation in all four quadrants. A standard 24-2 test grid for a right eye is shown in grey for reference.

This experiment was conducted in a similar manner to experiment one, utilising the same two-phase protocol as described in section *3.2.2.3*. Stimuli were presented at four of the eight locations in any one test, either the four locations at 22.8°, or the four locations at 36.2° from fixation. As these locations were equidistant from fixation, attentional bias should be largely eliminated. The participant completed one test for each of the five Goldmann stimuli with each of the test eccentricities, i.e. ten tests in total.

As there were only four locations in each test, phases were of a slightly shorter duration than in experiment one. The short MOCS phase was of approximately seven minutes duration, consisting of 180 presentations, with a rest break after 90. The standard MOCS phase was of approximately 30 minutes duration, consisting of 640

presentations, with a rest break after every 160. Again, the participant could request additional rest breaks as required.

A FOS curve was again constructed from the observer responses for each of the test locations, and fitted with a cumulative Gaussian psychometric function. Guess and lapse rates were permitted to vary.

To examine the association between sensitivity and visual field quadrant, p(seen, 0.5) values for each Goldmann stimulus were plotted against test location. This was also conducted with response variability values, to examine the association between response variability and visual field quadrant.

3.2.4 Results – experiment one

3.2.4.1 Psychometric functions

Examples of the stimulus levels used in the short MOCS phase in experiment one are shown for two test locations in *Table 3.2*; one location is for a Goldmann I stimulus, and one is for a Goldmann III stimulus, from two different participants. Observer responses, fitted with a cumulative Gaussian psychometric function, are shown for both test locations in *Figure 3.7*. 'Sens' indicates the sensitivity, and 'spread' indicates the response variability. P(seen) = 0.1 and 0.9 are indicated by the magenta dotted lines, and p(seen) = 0.3, 0.5, and 0.7 are indicated by the blue dotted lines; these stimulus levels were used in the standard MOCS phase.

Stimulus	Location	-9 dB	-4.5 dB	-3 dB	-1.5 dB	Estimated Sensitivity	+1.5 dB	+3 dB	+4.5 dB	+9 dB
G I	(63, -33)	13	17.5	19	20.5	22	23.5	25	26.5	31
G III	(3, -3)	0	4.5	6	7.5	9	10.5	12	13.5	18

Table 3.2 – Examples of pre-selected stimulus levels (dB) for two test locations, one for a Goldmann I, and one for a Goldmann III stimulus, for the short MOCS phase.
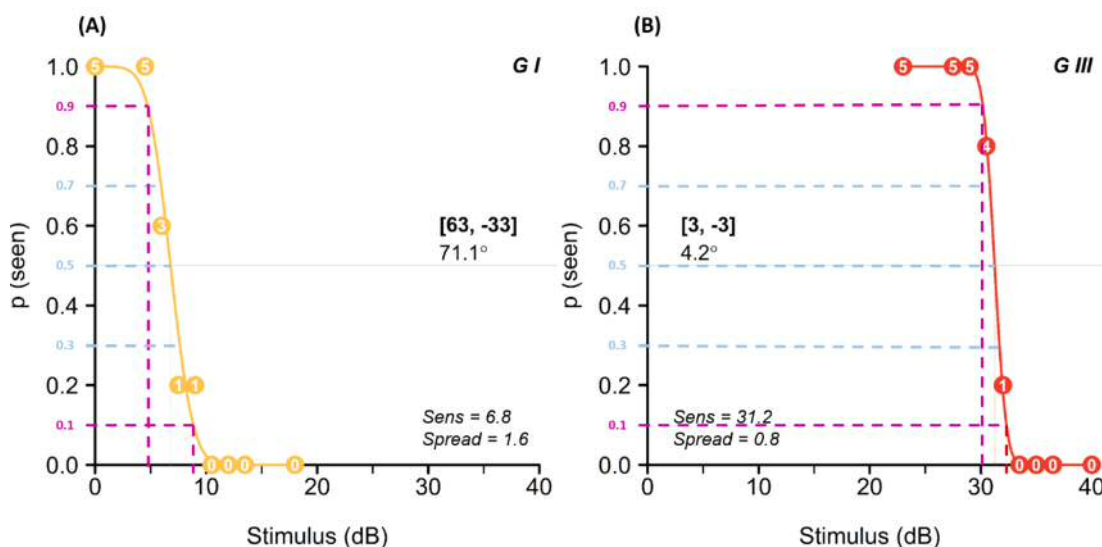
Figure 3.7 – Example FOS curves, fitted with a cumulative Gaussian psychometric function, for the short MOCS phase, for (A) a Goldmann I stimulus at [63,-33], and (B) a Goldmann III stimulus at [3,-3]. 'Sens' indicates the sensitivity, i.e. p(seen) = 0.5, and 'spread' indicates the SD of the psychometric function.

The step interval, and stimulus values used in the standard MOCS phase for the two test locations in *Table 3.2* and *Figure 3.7*, are shown in *Table 3.3*. Observer responses, fitted with a cumulative Gaussian psychometric function, are shown in *Figure 3.8*. Although the thresholds were quite different in the two examples, as one stimulus had a much smaller area than the other, and one test location was much further from fixation than the other, the response variability, i.e. the spread (SD) of the psychometric functions, were similar.

Stimulus	Location	Step	-15 dB	-5/-7 Step	-2 Step	-1 Step	P(seen)=0.5	+1 Step	+2 Step	+5/+7 Step
G I	(63, -33)	0.63	0.00	3.64	5.53	6.16	6.79	7.42	8.05	9.94
G III	(3, -3)	0.50	16.25	27.75	30.25	30.75	31.25	31.75	32.25	34.75

Table 3.3 – Examples of pre-selected stimulus levels (dB) for two test locations, one for a Goldmann I, and one for a Goldmann III stimulus, for the standard MOCS phase. 'Step' denotes the stimulus interval.
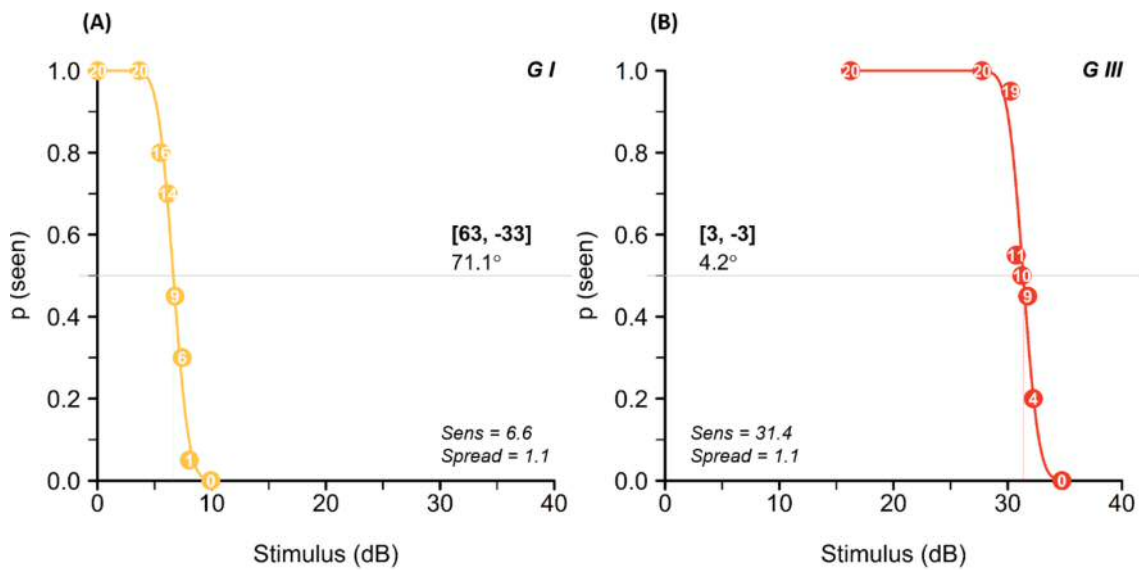
Figure 3.8 – Example FOS curves from the standard MOCS phase, for (A) a Goldmann I stimulus at [63, -33], and (B) a Goldmann III stimulus at [3, -3]. These are the stimulus levels and observer responses from the example short MOCS phase shown in Figure 3.7. 'Sens' indicates the sensitivity, i.e. p(seen) = 0.5, and 'spread' indicates the SD of the psychometric function.

Mean guess rate was 0.01, and mean lapse rate was 0.03. An example of the psychometric functions from the standard MOCS phase at one test location in experiment one, with all five Goldmann stimuli, is shown in *Figure 3.9*. The shapes of the psychometric functions are similar between the five Goldmann stimuli, suggesting a similar response variability between the five stimuli at this test location. This is a typical example of the psychometric functions obtained at all test locations, and with all Goldmann stimuli.
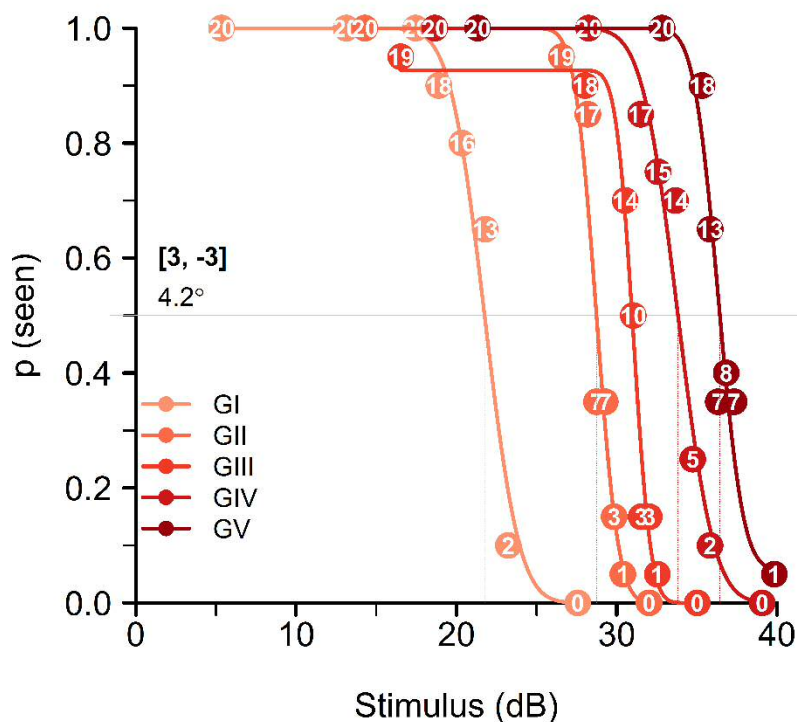
Figure 3.9 – Example of the five psychometric functions obtained for each of the five Goldmann stimuli at one test location.

One test location, [-33, 15], was the exception, with notable differences in the shape of the psychometric function with differing Goldmann stimuli for some participants. For one participant, a greater response variability was found with Goldmann III and IV in comparison with that obtained for Goldmann I, II and V (*Figure 3.10.A*). Response variabilities at the other four test locations were similar between all Goldmann stimuli, in keeping with the example shown in *Figure 3.9*. Repetition of the two-phase protocol on two separate days for Goldmann III and IV resulted in similar response variabilities to the other Goldmann stimuli (*Figure 3.10.B*), in keeping with findings for other participants and other test locations.

For another participant, a greater response variability was found with Goldmann I, II and III, in comparison with that obtained for Goldmann IV and V (*Figure 3.11.A*). As per the participant described above, response variabilities at the other four test locations were similar between all Goldmann stimuli. Repetition of the two-phase protocol on three separate days for Goldmann I, II and III, resulted in similar response

variabilities to the other Goldmann stimuli (*Figure 3.11.B*), in keeping with findings for other participants and all other test locations.
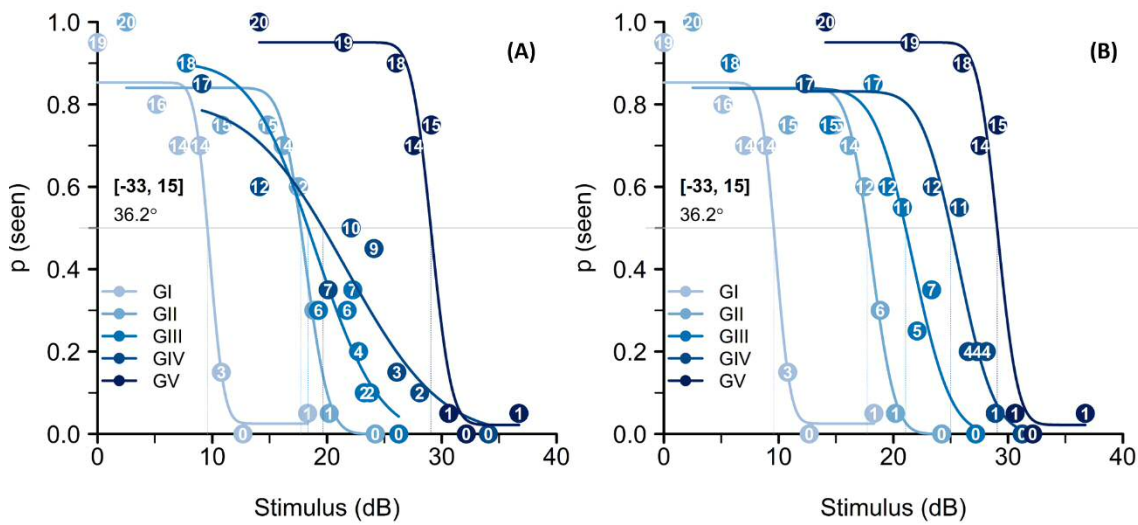
Figure 3.10 – Example FOS curves for the five Goldmann stimuli, for one participant at [-33,15]. (A) A greater response variability is noted with Goldmann III and IV in comparison with the other Goldmann stimuli. (B) When repeated, psychometric functions were more similar between all Goldmann stimuli, and more in keeping with those from other test locations.
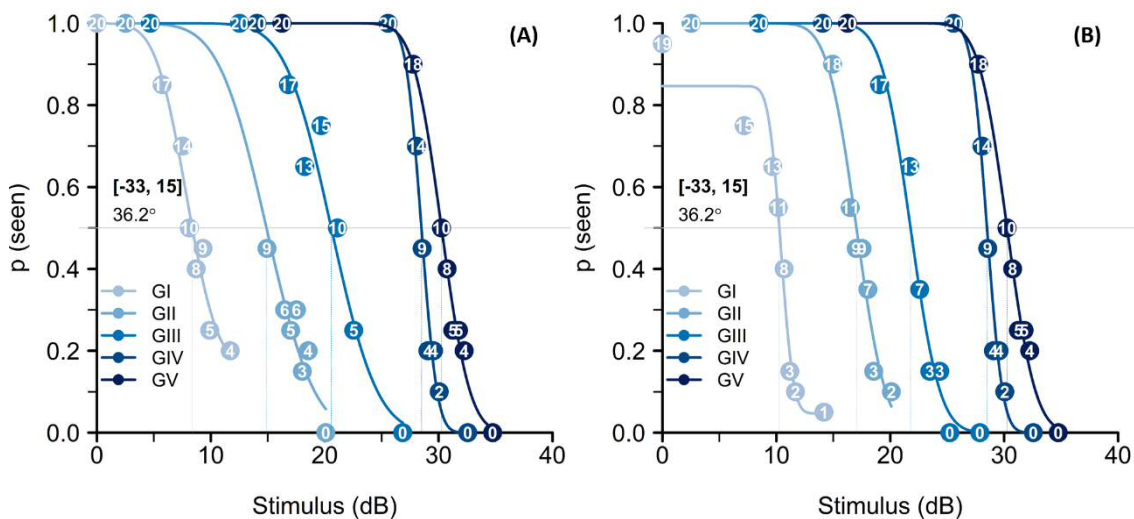
Figure 3.11 – Example FOS curves for the five Goldmann stimuli, for one participant at [-33,15]. (A) A greater response variability is noted with Goldmann I, II and III, in comparison with Goldmann IV and V. (B) When repeated, psychometric functions were more similar between all Goldmann stimuli, and more in keeping with those from other test locations.

It should be noted that the psychometric functions displayed for Goldmann I, II and III in *Figure 3.11.B* were obtained as part of experiment two. In this experiment, as described in section *3.2.3*, four equidistant locations were tested in all four quadrants; as such, the attentional bias differed in comparison to that of experiment one, in which five, non-equidistant locations were tested. Therefore, values from these psychometric functions were not used in any statistical tests, instead using values from psychometric functions shown in *Figure 3.11.A*.

3.2.4.2 Sensitivity

It was possible to obtain a sensitivity, p(seen) = 0.5, at all 125 test locations. *Figure 3.12* shows the mean sensitivity (dB) ± one SD at each test location, plotted as eccentricity from fixation, for Goldmann I-V. The lighter coloured background demarcates those test locations in the supero-nasal quadrant (22.8° [-21,9] and 36.2° [-33,15]) from those located in the infero-temporal quadrant.

A lower sensitivity was generally observed with increasing eccentricity from fixation for all Goldmann stimuli (*Figure 3.12*), although this relationship was not strictly linear. A greater sensitivity was observed at 44.3° (infero-temporal quadrant) in comparison to that at 36.2° (supero-nasal quadrant).

There appears to be a greater inter-individual variability between threshold sensitivities for smaller Goldmann stimuli, compared with larger stimuli at all test locations.
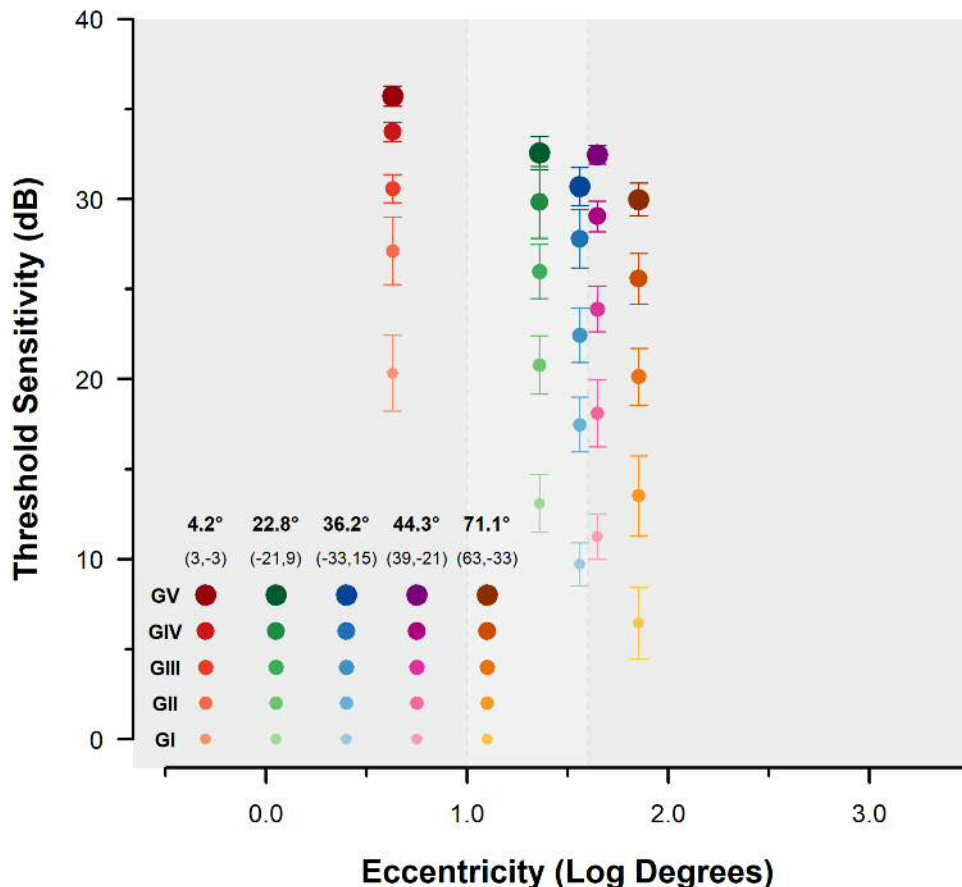
Figure 3.12 – Mean sensitivity (dB) plotted against eccentricity from fixation. Error bars indicate ± one SD, for Goldmann I-V. The lighter coloured background demarcates the supero-nasal test locations, from the infero-temporal (darker background) test locations.

3.2.4.3 Response variability

It was possible to obtain a response variability value at 124 of the 125 test locations; at one test location, the 75th percentile could not be established, and as such SD could not be ascertained. *Figure 3.13* shows the response variability ± one SD at each test location, plotted as eccentricity from fixation, for Goldmann I-V. Horizontal jitter has been added to aid data visualisation. As in *Figure 3.12*, the lighter coloured background demarcates those test locations in the supero-nasal quadrant (22.8° [-21,9] and 36.2° [-33,15]) from those located in the infero-temporal quadrant.

From *Figure 3.13*, it is difficult to observe any definitive association between response variability and eccentricity. The smaller Goldmann stimuli generally appear to have a greater response variability than the larger Goldmann stimuli, however there is

substantial overlap in response variability values between Goldmann stimuli (error bars shown in *Figure 3.13*). Additionally, there appears to be a greater inter-individual variability in response variability with smaller Goldmann stimuli compared with larger Goldmann stimuli.

No statistically significant differences were found between response variabilities for the five Goldmann stimuli at any test location (p > 0.05 for all comparisons).
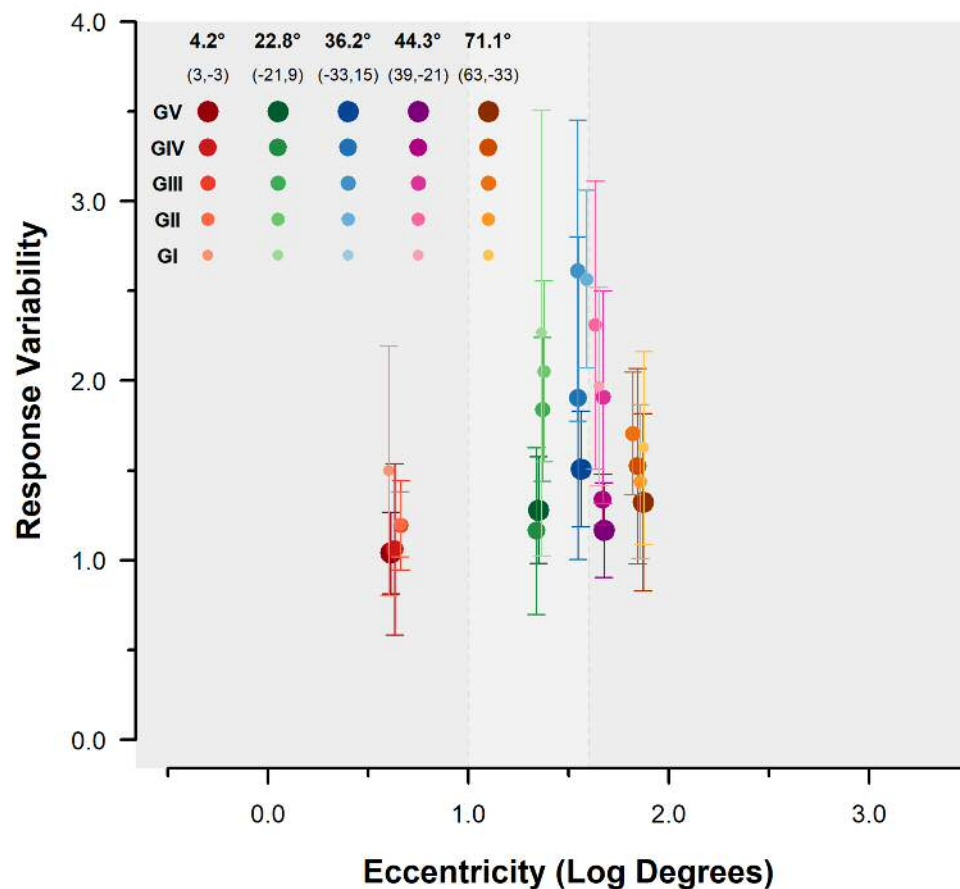
Figure 3.13 – Mean response variability, plotted against eccentricity from fixation, with error bars indicating ± one SD, for Goldmann I-V. The lighter coloured background demarcates the supero-nasal test locations, from the infero-temporal (darker background) test locations. Horizontal jitter has been added to aid data visualisation.

3.2.5   Results – experiment two

Example FOS curves from one test location at 22.8°, and one at 36.2°, in experiment two are shown in *Figure 3.14*, for all Goldmann stimuli. As noted in experiment one (section *3.2.4*), the shape of the psychometric functions is similar between the five Goldmann stimuli, suggesting a similar response variability between Goldmann stimuli at these test locations. These are typical examples of the psychometric functions obtained at all locations tested. Mean guess rate was 0.01, and mean lapse rate was 0.02.
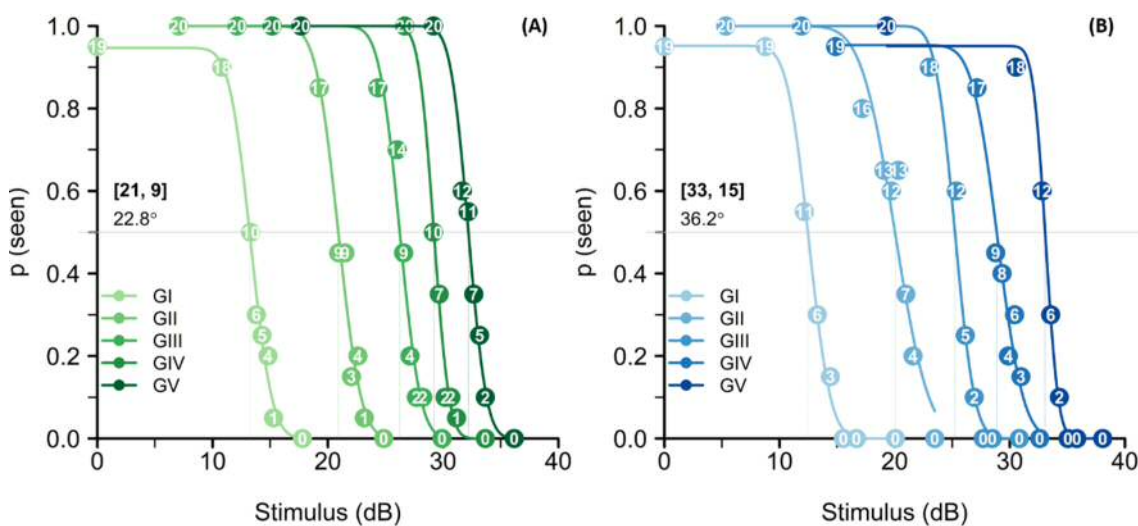
Figure 3.14 – Examples of FOS curves obtained for one test location at 22.8°, and one at 36.2°, from experiment two.

Sensitivity

It was possible to obtain a sensitivity (p(seen) = 0.5) at all 40 test locations. *Figure 3.15* shows the threshold sensitivities with each of the five Goldmann stimuli for the two eccentricities, 22.8° (*Figure 3.15.A*) and 36.2° (*Figure 3.15.B*) from fixation, in each of the four visual field quadrants. Threshold sensitivities for the nasal locations are displayed to the left, on the lighter background, and those for the temporal locations are displayed to the right, on the darker background. The outer data points denote inferior locations, and the inner denote the superior locations. Threshold sensitivities appeared largely similar at all four test locations for both eccentricities, although

appeared slightly lower at superior compared with inferior locations at both eccentricities. Nasal and temporal sensitivities were conflicting between the two eccentricities, appearing slightly higher at nasal compared with temporal locations at 22.8° (*Figure 3.15.A*), and, to a slightly greater extent, lower at nasal compared with temporal locations at 36.2° from fixation (*Figure 3.15.B*).
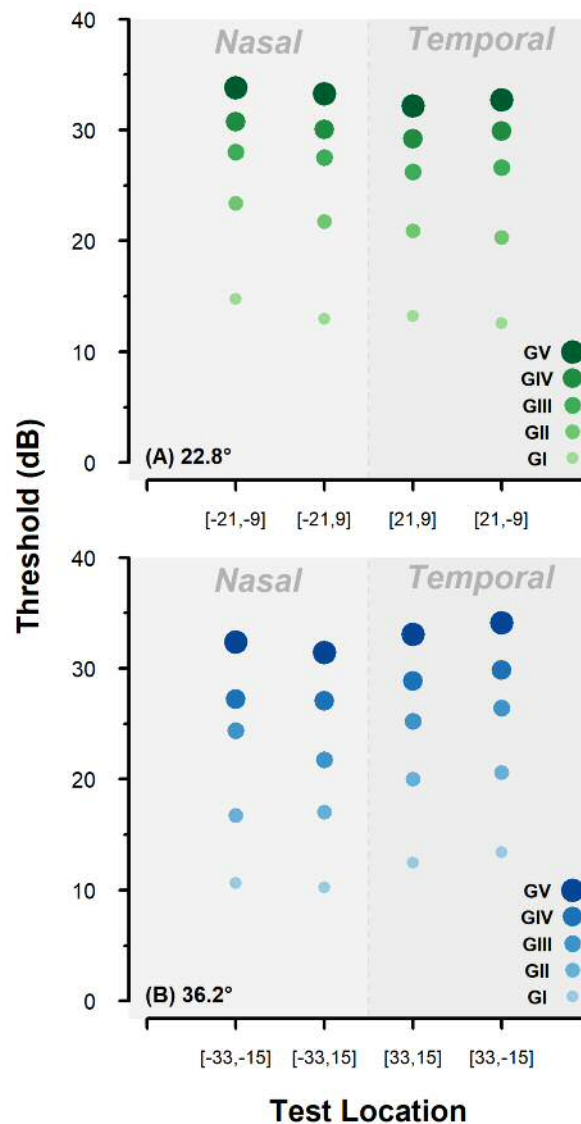
Figure 3.15 – Sensitivity for each of the eight locations tested, at (A) 22.8°, and (B) 36.2° from fixation, for each of the five Goldmann stimuli.

Response variability

It was possible to establish a response variability at all 40 locations tested. *Figure 3.16* shows the response variabilities with each of the five Goldmann stimuli for the two eccentricities, 22.8° (*Figure 3.16.A*) and 36.2° (*Figure 3.16.B*) from fixation, in each of the four visual field quadrants. Horizontal jitter has been added to aid data visualisation. As per *Figure 3.15*, response variabilities for the nasal locations are displayed to the left, on the lighter background, and those for the temporal locations are displayed on the right, on the darker background. The outer data points denote inferior locations, and the inner denote the superior locations. Response variability appears largely similar in all four quadrants, and between the two eccentricities tested; any differences between test locations are not consistent per hemifield. Generally, larger stimuli appear to show a lower response variability than smaller stimuli, but this is inconsistent across the eight test locations.
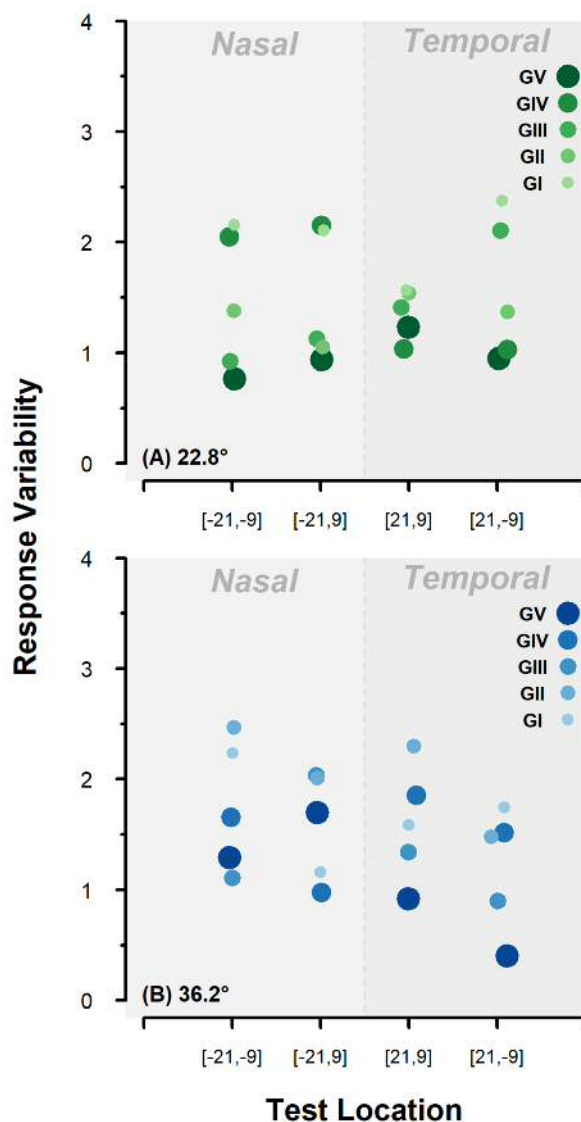
Figure 3.16 – Response variability for each of the eight locations tested, at (A) 22.8°, and (B) 36.2° from fixation, for each of the five Goldmann stimuli. Horizontal jitter has been added to aid data visualisation.

3.2.6   Discussion

These two small experiments investigated whether the previously reported reduction in response variability with the use of larger perimetric stimuli (Wall et al. 1997) could be artefactual, due to sub-optimal stimulus selection for a MOCS procedure. The findings of this study suggest this may be the case. Although some differences were observed between response variability measures for different Goldmann stimuli (*Figure 3.13*), these were not statistically significant, and are unlikely to be clinically significant. Response variability values appeared largely similar between the five

stimuli independent of eccentricity. This may also suggest that response variability is largely independent of sensitivity, although it would be necessary to perform this experiment in the presence of disease to confirm this.

This highlights the importance of careful stimulus pre-selection when utilising a MOCS procedure, in order to minimise bias and to ensure that any differences found may be attributed to the stimulus configuration itself.

Differences between response variabilities with Goldmann stimuli were noted at 36.2° [-33,15] for two participants, in which a shallower psychometric function was observed with Goldmann III and IV compared with Goldmann I, II, and V for one participant (*Figure 3.10*), and a shallower psychometric function was observed with Goldmann I, II, and III compared with Goldmann IV and V for another participant (*Figure 3.11*). However, this was not found to be repeatable, and psychometric function shape was similar at this test location to those at the other four test locations when the MOCS procedure was repeated with these Goldmann stimuli. Heijl et al. (1987) noted that mean perimetric sensitivities decreased more rapidly in the nasal visual field, and particularly in the superior quadrant, which may support this finding of a greater response variability in this quadrant. In addition, it has been reported that the most reliable data is obtained in kinetic perimetry when approaching the isopter perpendicularly (Lynn et al. 1991); as the hill of vision is steeper in the nasal visual field, it is possible this results in an increase in variability, although one would perhaps expect a repeatable psychometric function shape. Redmond and Artes (2012), who conducted the pilot studies prior to commencement of this PhD, reported similar findings at this same test location. Contrary to the experiment presented here however, consistently shallower psychometric functions were observed on repeat testing (personal communication).

It is possible that this finding indicates an attentional 'bias' caused by the visual field locations chosen; two locations were in the supero-nasal quadrant, and three were in the infero-temporal quadrant, and all test locations were at differing eccentricities from fixation.

As expected, a reduction in sensitivity was observed with increasing visual field eccentricity with all Goldmann stimuli, although this relationship was not strictly linear. In *Figure 3.12*, a greater sensitivity is observed at 44.3° in comparison with that at 36.2°, despite the increase in eccentricity from fixation. It is likely that this represents the known differences in sensitivity between visual field quadrants, as it has long been observed that visual field sensitivity is not symmetrical between nasal and temporal, nor superior and inferior quadrants. Isopter plots from kinetic, manual perimetry have demonstrated that isopters extend further, indicating a greater sensitivity, in the inferior compared with the superior field, and, more notably, extend further in the temporal compared with the nasal field (*Figure 3.17*), a difference which is not solely attributable to limits imposed by the position of the nose (Traquair 1927; Tate and Lynn 1977).
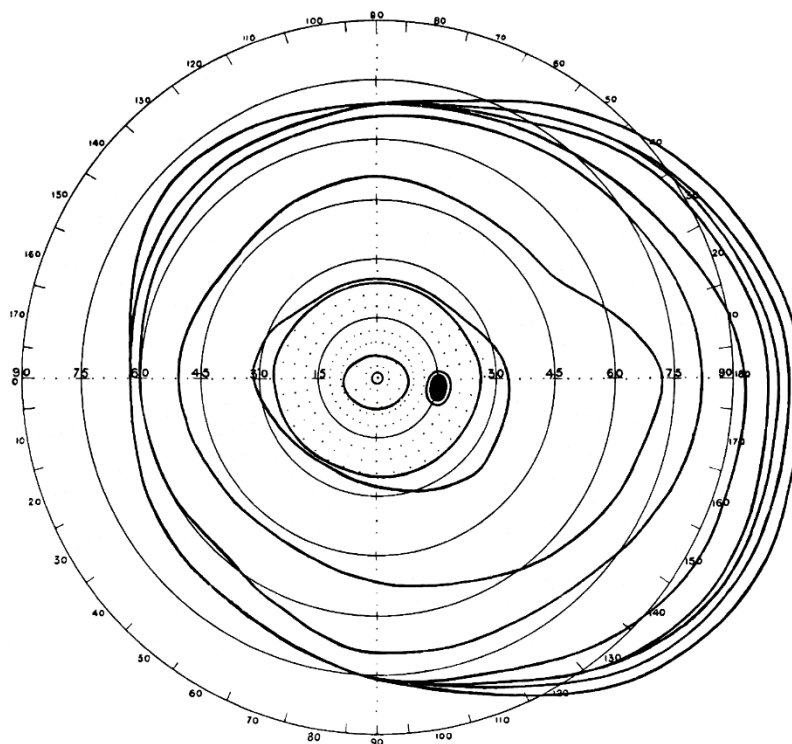
Figure 3.17 – Isopter plots for perimetric stimuli of differing area. The inner isopter is plotted for the smallest stimulus, with increase in isopter size with increase in stimulus area (Traquair 1927).

This may be further examined in *Figure 3.15*, in which small differences in sensitivity can be observed at test locations in all four quadrants, equidistant from fixation. Sensitivity was higher at inferior test locations compared with superior at both eccentricities, in keeping with isopter plots as described above, although slight differences were noted between eccentricities when comparing nasal and temporal locations. At 22.8° from fixation, temporal locations demonstrated a lower sensitivity than nasal locations, while at 36.2°, temporal locations demonstrated a higher sensitivity. This may be attributed to the 'visual streak', an area of higher retinal ganglion cell density in the nasal retina (i.e. the temporal visual field) along the horizontal meridian, demonstrated anatomically by Curcio et al. (1990), and psychophysically by Anderson et al. (1992). Curcio et al. (1990) noted that differences between nasal and temporal retina became more pronounced with increasing eccentricity from the fovea. Observation of the isopter plots in *Figure 3.17* demonstrates a similar sensitivity between nasal and temporal fields up to 30° from fixation, with increasing asymmetry between nasal and temporal fields beyond this eccentricity. This may explain why temporal test locations appeared to have a lower sensitivity at 22.8° from fixation, but a higher sensitivity at 36.2° from fixation (*Figure 3.15*). It is possible that the reduced sensitivity in the nasal field, coupled with the higher response variability observed with some observers, may indicate an underlying weakness of the visual field in this area, which could be susceptible to damage in the presence of glaucoma. This may explain why a nasal step is often an early feature of a glaucomatous visual field (Drance 1969).

The use of the two-phase MOCS procedure and the use of non-integer dB values reduces some of the potential bias associated with the use of integer values on the logarithmic dB scale, however it does not eliminate this bias completely. Due to differing stimulus area, a dB modulation with a small stimulus represents a smaller raw energy increment than a dB modulation with a larger stimulus, irrespective of its position on the dB scale. As such, the comparisons between Goldmann stimuli presented here are not direct, and caution should still be exercised when attempting to draw firm conclusions from the findings presented here.

3.3   Choice of psychometric function

In the experiments described in section *3.2*, the FOS curves were fitted with a cumulative Gaussian psychometric function, however there are numerous types of psychometric model available. The three most commonly used psychometric function models are the cumulative Gaussian, the logistic, and the Weibull (Harvey 1986; Hill 2001; Gilchrist et al. 2005). A brief description of each follows.

1. Cumulative Gaussian model: sometimes referred to as a 'probit' model, the cumulative Gaussian model is the cumulative distribution function of a normal distribution, thus making the assumption that the underlying data error is normally distributed (Treutwein and Strasburger 1999; Kingdom and Prins 2009).

2. Logistic model: this model is the cumulative distribution function of a logistic distribution, thus making the assumption that the underlying data error is logistically distributed (Kingdom and Prins 2009). This is almost identical in shape to a cumulative Gaussian distribution, although a logistic distribution has slightly heavier tails (Wetherill and Levitt 1965). As this model is computationally more efficient, with a closed analytic form, it may be used as a substitute for a cumulative Gaussian model (Berkson 1951; Harvey 1986).

3. Weibull model: this model makes the assumption that the underlying data error is a Weibull distribution, which often displays either a negative or positive skew. However, this distribution can be quite flexible, such that it may approximate a logistic or cumulative Gaussian distribution (Rinne 2008). Care should be taken when fitting this model to linear data, as the threshold (α), and slope (β) are not independent of each other, however this is not the case when fitting to logarithmic data, known as a Gumbel model (Kingdom and Prins 2009; May and Solomon 2013). The Weibull model is used in the QUEST algorithm, and although it is based on the older theory of 'high-threshold detection', which, amongst other factors, assumes near-perfect responses from the observer, it is noted to provide a good fit to FOS data (Watson and Pelli 1983; May and Solomon 2013).

Most other types of psychometric function are a modification of one of these three models. While knowledge of the theoretical behaviour of a psychometric function model may be used to select one in favour of another, there may be little difference between models in practice, as they adopt structurally similar shapes, although a goodness-of-fit analysis, establishing which psychometric function best fits the collected data, may inform this decision (Harvey 1986; Hill 2001; Gilchrist et al. 2005).

In the experiment presented in chapter four, a MOCS procedure, with a similar multi-phase protocol as described in section *3.2.2.3*, was utilised. During this experiment, the FOS curves obtained in both phases were fitted with a cumulative Gaussian psychometric function, using MATLAB (version R2014b; The MathWorks Inc., Natick, MA, USA) and the modelfree toolbox (Zychaluk and Foster 2009). Subsequent to data collection, and prior to statistical analysis of differences between threshold and response variability with differing stimulus forms, it was necessary to ensure that the selected psychometric model provided the most appropriate fit of the resulting FOS curves, as an inappropriate selection could lead to inaccurate conclusions with respect to differences between stimulus forms. To inform the psychometric function selection, all FOS curves obtained as part of the experiment described in chapter four were fitted with the three most common psychometric functions, and a goodness-of-fit analysis conducted.

3.3.1.1 Goodness-of-fit

In the experiment described in chapter four, 800 locations were tested in total. Three psychometric functions, cumulative Gaussian, logistic and Gumbel (i.e. log-Weibull, recommended as an alternative to the Weibull when fitting logarithmic data), were fitted to each of these 800 test locations, using MATLAB (version R2015b; The MathWorks Inc., Natick, MA, USA) and the Palamedes toolbox (Prins and Kingdom 2009). This toolbox permitted the guess and lapse rates to vary, which may reduce threshold and slope bias (Treutwein and Strasburger 1999; Wichmann and Hill 2001a). Three different methods were used to fit the psychometric function with each of the three models as follows:

1. The guess and lapse rates were fixed at zero.

2. Both guess and lapse rates were permitted to vary between 0 and 0.1 (0-10%), such that the psychometric function model estimated the most appropriate guess and lapse rates within these limits from the available data.

3. The guess rate was fixed at zero, and the lapse rate was permitted to vary between 0 and 0.1 (0-10%). As the experiment described in chapter four included participants with glaucoma, in whom a complete FOS curve was not possible due to a reduction in sensitivity, the lapse rate will be affected, although the guess rate may not.

For each model fit, the residuals were calculated for each of the data points from the psychometric functions (6400 residuals in total) as per *Figure 3.18*; residuals were negative for data points falling below the psychometric function, and positive for data points located above the psychometric function. Therefore, the psychometric function which more closely fitted the data would have residuals with a mean close to 0, and a minimal SD.

Histograms of the residuals for the three psychometric function models are shown in *Figure 3.19*, for guess/lapse rates permitted to vary between 0 and 0.1, and guess/lapse rates fixed at 0. Mean and SD of the residuals are displayed on each plot.
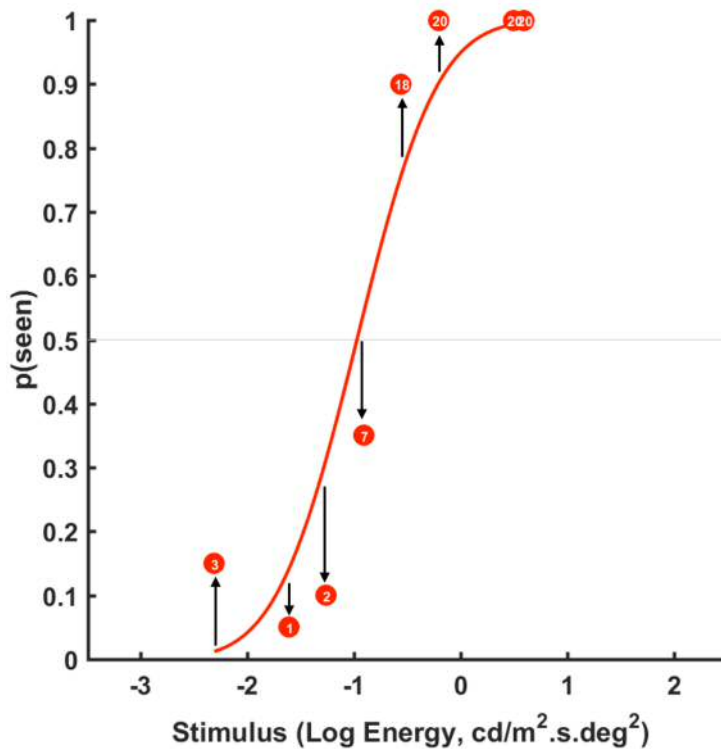
Figure 3.18 – Example of a psychometric function, with residuals indicated by the black arrows.

From *Figure 3.19*, it is apparent that the logistic psychometric function provided a closer fit of the data, with mean residuals closest to 0, both when guess and lapse rates were fixed at 0, and when they were permitted to vary between 0 and 0.1. The logistic psychometric function also had the smallest SD, i.e. the least spread of the residuals. Permitting guess and lapse rates to vary between 0 and 0.1 demonstrated a closer fit of the data than fixing these rates at 0.

Although varying guess and lapse rates between 0 and 0.1 generally provided an overall closer fit of the data (an example of which is shown in *Figure 3.20*), observation of the individual psychometric functions revealed that was not true for all FOS curves. Some FOS data were fitted more closely with a psychometric function in which guess and lapse rates were fixed at 0 (e.g. *Figure 3.21*), while in others, fixing the guess rate at 0, and permitting the lapse rate to vary between 0 and 0.1, provided a more representative fit of the dataset (e.g. *Figure 3.22*).
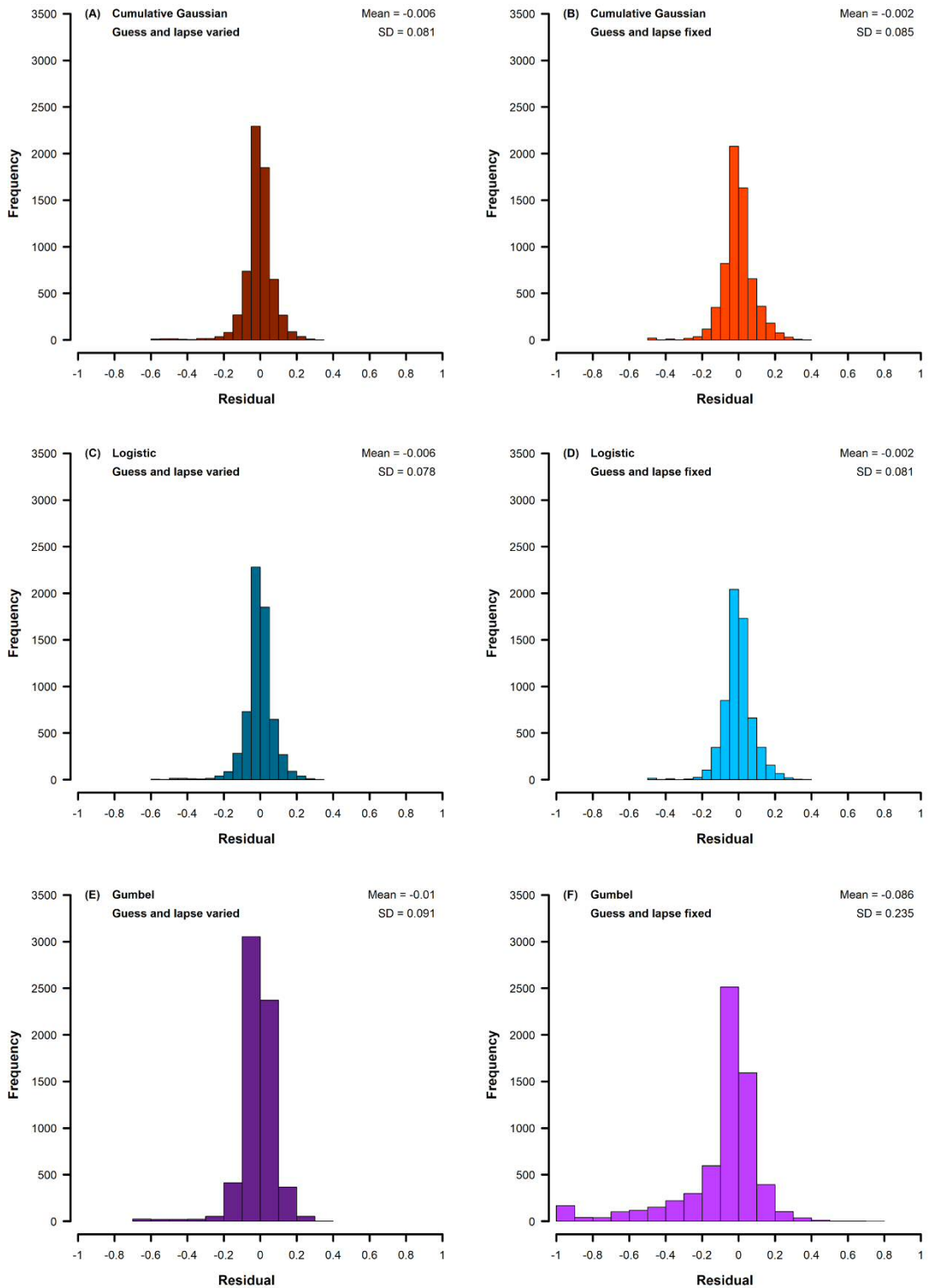
Figure 3.19 – Residuals from the goodness-of-fit analysis, for three psychometric functions (A) & (B) cumulative Gaussian, (C) & (D) logistic, and (E) & (F) Gumbel. On the left, guess and lapse rates have been permitted to vary between 0 and 0.1, and on the right, guess and lapse rates have been fixed at 0.
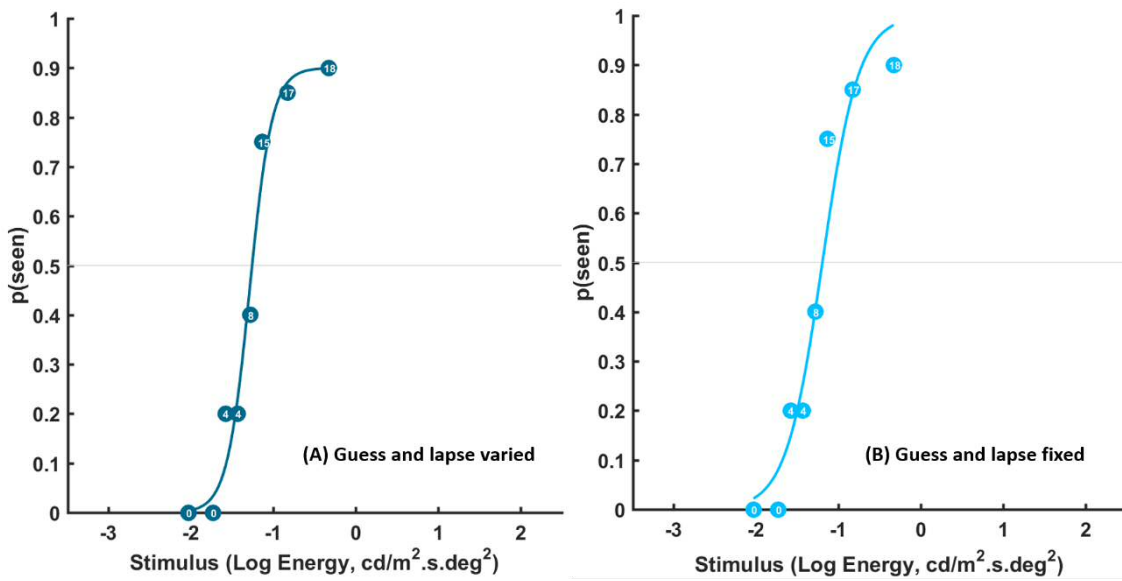
Figure 3.20 – Example FOS data in which permitting guess and lapse rate to vary between 0 and 0.1 permits a closer fit of the psychometric function.
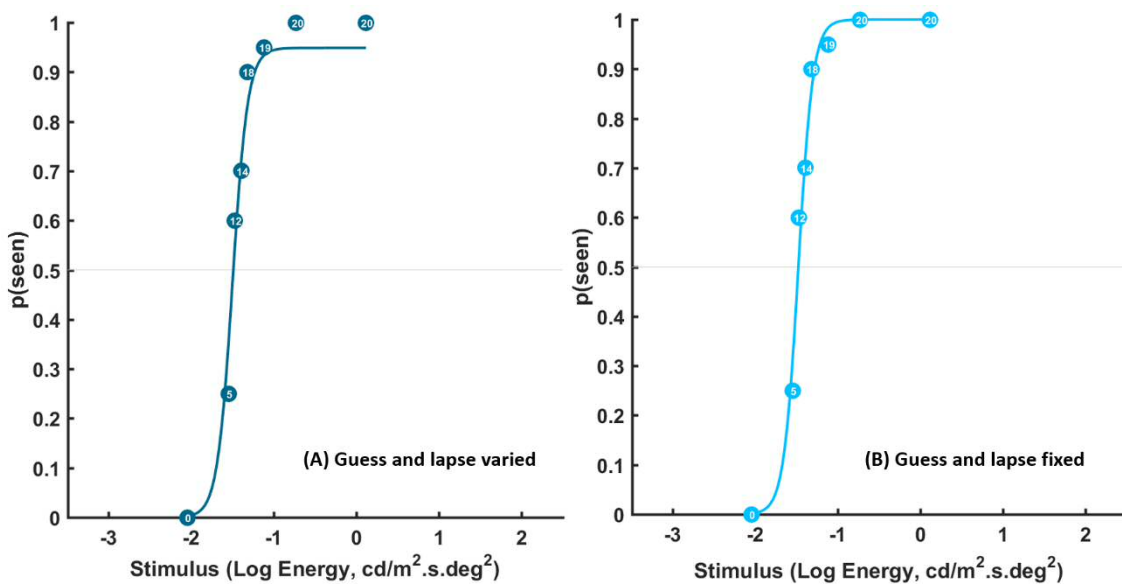
Figure 3.21 – Example FOS data in which fixing guess and lapse rates at 0 permits a closer fit of the psychometric function.
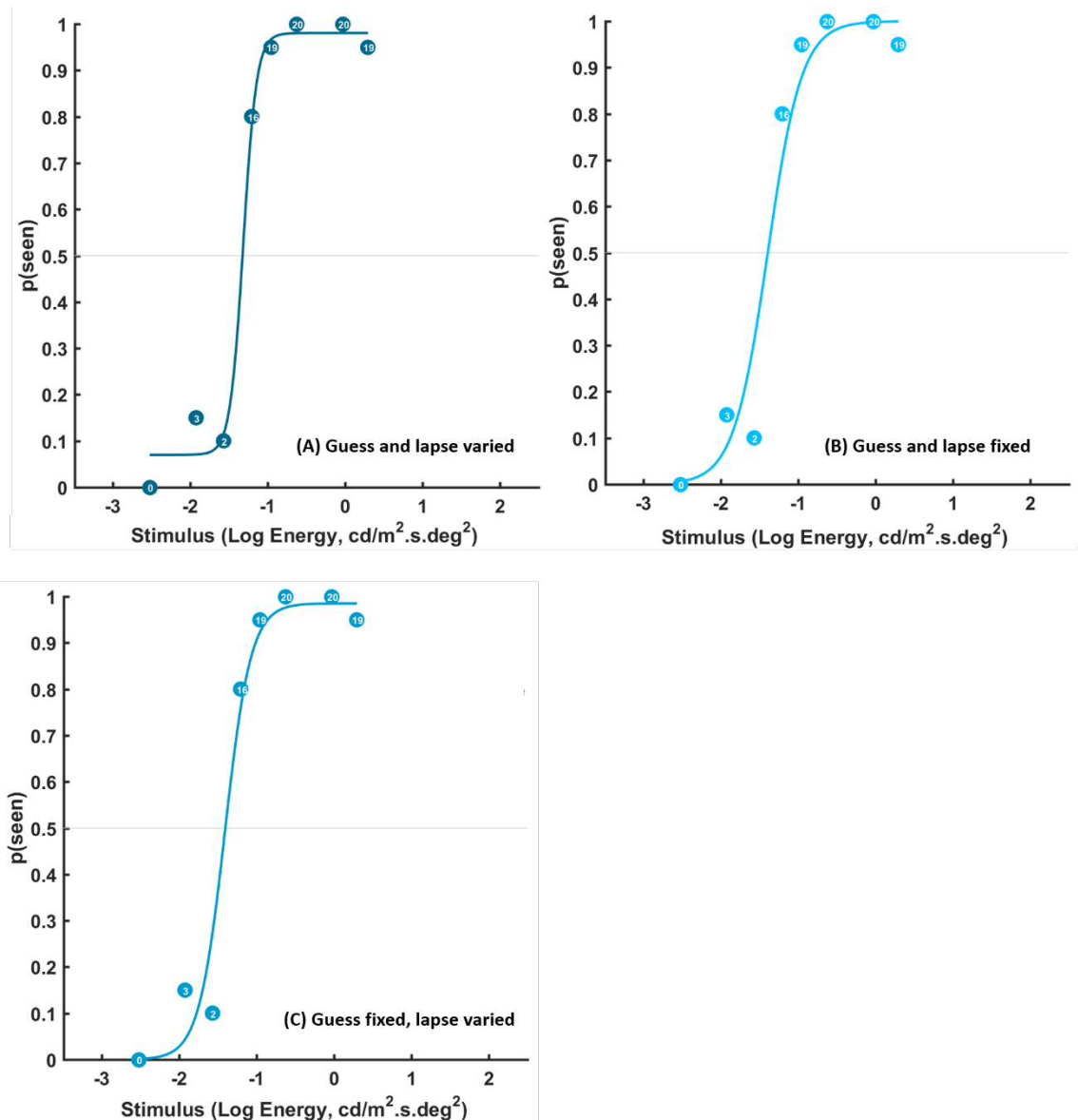
Figure 3.22 – Example FOS data in which fixing guess rate at 0, while allowing lapse rate to vary between 0 and 0.1, permits a closer fit of the psychometric function.

3.4   Considerations for subsequent experiments

This chapter details some of the preliminary investigations undertaken in preparation for the design, execution, and statistical analysis of the experiment described in chapter four. Based on the findings presented here, a number of considerations were applied to this subsequent experiment, as detailed below.

3.4.1 Test locations

Given the uncertainties that may arise regarding attentional bias and the impact this could have on threshold and response variability results, as detailed in section *3.2.6*, it seemed prudent to control for this in the experiment described in chapter four. As such, only four visual field locations were tested, positioned equidistant to fixation as in experiment two presented here. As an added precaution, locations were selected along the 45°, 135°, 215° and 315° meridians, such that the horizontal and vertical distances from fixation were equal, similar to the locations used by Redmond et al. (2010a; 2010b).

3.4.2 Stimulus design

In the experiment presented in chapter four, stimuli modulated in both area and luminance increments, which are not easily comparable. To enable a direct comparison between these stimuli, a common scale must be used. This can be achieved by determining the energy of the stimulus (*Equation 3.3*), as per Mulholland et al. (2015b; 2015c).

$$E = L \times A \times D$$

Equation 3.3

$E = Energy\ (cd/m^2.s.deg^2)$

$L = Luminance\ (cd/m^2)$

$A = Area\ (deg^2)$

$D = Duration\ (s)$

As discussed in section *3.2.6*, the use of a common scale alone does not enable a direct comparison to be made between different stimuli. Therefore, stimulus forms used in chapter four were designed such that increment steps were approximately equal in terms of stimulus energy, across different stimuli, irrespective of whether that stimulus modulated in area, contrast, or both simultaneously.

3.4.3 Psychometric function fitting

Based on the findings of section *3.3*, all FOS data collected in chapter four as part of the standard MOCS phase were fitted with a logistic psychometric function, using MATLAB (version R2015b; The MathWorks Inc., Natick, MA, USA) and the Palamedes toolbox (Prins and Kingdom 2009). As no single method of guess and lapse rate fitting was found to satisfactorily fit all FOS data at all test locations, each FOS curve was fitted with all three guess/lapse rate methods detailed in section *3.3.1.1*, and examined individually to determine which method provided a more representative fit of the FOS curve data. Threshold and response variability were then determined from the most appropriate psychometric function fit; as the logistic model is an acceptable substitute for a cumulative Gaussian model (Berkson 1951; Harvey 1986), SD of the psychometric function was established as per *Equation 3.2*, as an indication of response variability.

Chapter 4 Quantifying the signal/noise ratio with perimetric stimuli optimised to probe changing spatial summation in glaucoma

4.1   Introduction

As discussed in section *1.3.3*, while considered the current clinical standard for identifying glaucomatous visual field damage and change over time (NICE 2009; 2017), SAP has three cardinal limitations. First, SAP has poor sensitivity to early disease, and although test-retest variability is lowest in early disease and in healthy individuals (Tafreshi et al. 2009), it is unacceptably high for the identification of subtle damage (Wilensky and Joondeph 1984; Artes et al. 2002a). Second, the greater variability in visual field locations with moderate damage (which increases with depth of defect) greatly inhibits the timely identification of change in those with established glaucoma (Heijl et al. 1989a; Wall et al. 1996; Artes et al. 2002a). Third, the test has a limited useable dynamic range, with test-retest variability spanning almost its entire range in individuals with advanced damage, such that the measurement of remaining vision is difficult (Wilensky et al. 1986; Artes et al. 2002a; Gardiner et al. 2014).

Several studies have attempted to address the limitations of SAP by investigating the utility of alternative stimuli and comparing it to that of the clinical standard (Goldmann III). It has been suggested that employment of some alternative stimuli (e.g. the larger Goldmann V stimulus, area: 2.3 deg^2) could enable measurement of a larger range of damage, with an accompanying reduction in test-retest variability (Wilensky et al. 1986; Wall et al. 1997). This addresses the 'noise' component of the SNR, but ascertaining whether such stimuli allow the test to maintain the same sensitivity to early damage ('signal') is not straightforward. In the absence of a clear rationale for using alternative stimuli, in terms of physiology, beyond reports that they may offer lower measurement variability, it is premature to confirm their superior utility, or otherwise, in clinical testing. Furthermore, a comparison of the utility of different stimuli on existing clinical instruments is not straightforward, particularly if it is not possible to precisely control their parameters, and without a precise knowledge of the workings of the threshold algorithm employed. A comparison of stimuli with a non-

clinical technique, such as MOCS, is also difficult if one is restricted to using the stimulus step size and scale provided on the clinical instrument. This becomes even more challenging if comparing stimuli between different instruments. Such a restriction could well affect the resolution and accuracy with which the psychometric functions can be sampled for different stimuli, thereby increasing the risk of slope bias (Hill 2001; Wichmann and Hill 2001a; Wichmann and Hill 2001b). A full understanding of the diagnostic benefits of using alternative stimuli requires the removal or minimisation of confounding factors that are unrelated to the stimulus configuration, such as the threshold algorithm or unequal psychometric function sampling between stimuli.

The optimisation of stimulus parameters for use in SAP to maximise SNR should be based on the underlying physiological mechanisms being measured. As discussed in section *1.4*, spatial summation describes the way in which the visual system integrates light energy across the area of a stimulus. Ricco's law states that, for a range of small stimuli, within a critical area (Ricco's area), the intensity of the stimulus at threshold is inversely proportional to its area (Ricco 1877), as per *Equation 1.2*. Ricco's area is not a constant value, and has been found to vary with visual field eccentricity (Wilson 1970; Volbrecht et al. 2000b; Vassilev et al. 2003; Khuu and Kalloniatis 2015a), retinal illuminance (Glezer 1965; Lelkens and Zuidema 1983; Redmond et al. 2013b), and stimulus duration (Wilson 1970; Scholtes and Bouman 1977). Traditionally, Ricco's area was thought to have a physiological basis at the retinal level (Glezer 1965; Ikeda and Wright 1972; Fischer 1973), however increasing evidence indicates that it is likely a perceptual result of spatial filtering at multiple hierarchies of visual processing, in the retina and at the visual cortex (Ransom-Hogg and Spillmann 1980; Schefrin et al. 1998; Vassilev et al. 2000; Anderson 2006; Pan and Swanson 2006; Je et al. 2018); i.e. the 'perceptive field' (Vassilev et al. 2005; Anderson 2006). An enlarged Ricco's area has been found in patients with POAG, and differential amounts of sensitivity loss to a range of stimulus areas can be mapped to a lateral shift in the spatial summation function (Redmond et al. 2010a). The finding has important implications, not only for a better understanding of the pathophysiological changes that occur in glaucoma, but also for the development of methods to identify early subtle damage (Anderson 2006;

Redmond et al. 2010a). Pan and Swanson (2006) have shown that, rather than probability summation across retinal ganglion cells, it is spatial filtering by multiple cortical mechanisms that accounts for perimetric spatial summation. Although glaucoma is characterised by ganglion cell death, it is perhaps unsurprising then that it is difficult to reconcile perimetric sensitivity and retinal structure, without consideration being given to spatial summation. Given the dependence of the relationship between visual field sensitivity (with conventional stimuli) and underlying ganglion cell density, on the relative area of the stimulus and local Ricco's area (Swanson et al. 2004), in addition to the variation in Ricco's area with visual field eccentricity (Graham and Bartlett 1939; Graham et al. 1939; Hallett 1963; Wilson 1970; Vassilev et al. 2003), and known changes in Ricco's area with glaucoma (Redmond et al. 2010a; Kalloniatis and Khuu 2016), it makes little sense to continue using an arbitrary fixed-area stimulus to probe the visual field. Rather, a stimulus should be selected in a way that supports meaningful measurements of changes in spatial summation in glaucoma. Anderson (2006) proposed that, if early glaucomatous loss were associated with a change in spatial summation in glaucoma, greater attention should be given to the area of the stimulus relative to Ricco's area; specifically, scaling the stimulus to the local Ricco's area in healthy individuals. Given that the intensity threshold at Ricco's area is largely constant irrespective of visual field locus (Wilson 1970), deviations from normal could be measured and compared on an equal par between locations. Redmond et al. (2010a), having found an enlarged Ricco's area in glaucoma, proposed a test paradigm whereby a relative shift in the spatial summation function in glaucoma might be better identified by varying stimulus area during the test, either instead of, or simultaneously with contrast. This is illustrated in *Figure 4.1*; the spatial summation curve to the left indicates a healthy observer, and the spatial summation curve to the right indicates an observer with glaucoma. Ricco's area, where the slope of the summation curve changes, is indicated on both curves. The arrow labelled 'GIII' indicates a Goldmann III stimulus, of fixed area varying in contrast, as is currently used in SAP; it can be seen that the use of such a stimulus identifies the difference between the curves at its narrowest point.
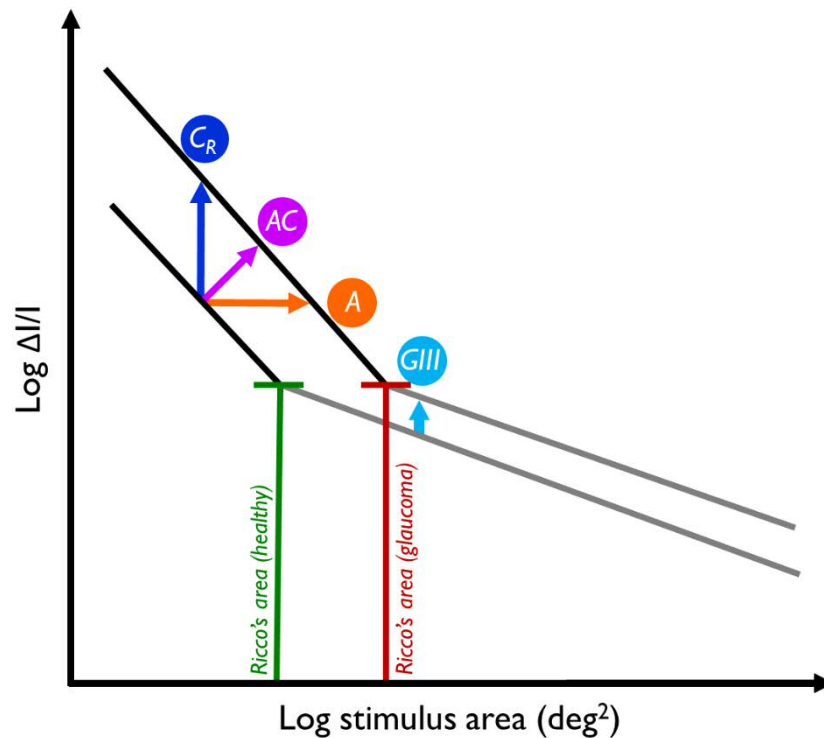
Figure 4.1 – Schematic spatial summation curves in patients with glaucoma and age-similar controls, obtained by measuring sensitivity for a range of stimulus areas (Rountree et al. 2018, adapted from Redmond et al. 2010a).

In the experiment presented here, the hypothesis of Redmond et al. (2010a) was tested, as illustrated in *Figure 4.1*, that a stimulus varying in area alone (A), or simultaneously with stimulus contrast (AC), will enable a greater disease signal by directly measuring a shift in an individual's spatial summation function. In addition, it is hypothesised that the use of such a stimulus, varying in area rather than contrast-only, will have reduced response variability compared to that found with conventional stimuli. In this experiment, the disease signal, response variability, and SNR for four different stimulus forms were compared, as illustrated in *Figure 4.1*, varying in area, contrast (two stimuli of different, but fixed area), and both area and contrast simultaneously.

4.2   Methods

In this cross-sectional study, psychometric functions were measured with four different stimulus forms (two varying in contrast only, one varying in area only, and

one varying proportionally in area and contrast simultaneously), in participants with glaucoma and age-similar controls. Disease signal (deviation in energy threshold from that of age-matched normal), noise (response variability), and SNR were determined and compared between stimulus forms.

4.2.1   Participants

Thirty participants with glaucoma (median [IQR] age: 70.5 years [66.5, 74.7]; median [IQR] MD: -4.04 dB [-9.30, -2.78]) and 20 age-similar healthy participants (median [IQR] age: 67.3 years [62.0, 75.1]; median [IQR] MD: +0.33 dB [-0.40, +0.77]) were recruited and tested. All of the glaucoma participants had received a diagnosis of POAG, 18 with high tension and 12 with normal tension glaucoma, by the hospital eye service. Glaucoma severity varied from minimal field loss ('within normal limits' on the Glaucoma Hemifield Test) to 'advanced' field loss (categorised with the Hodapp-Parrish-Anderson glaucoma grading scale; Hodapp et al. 1993), with the SITA Standard 24-2 program on the HFA II. All healthy participants had a full visual field ('within normal limits' on the Glaucoma Hemifield Test).

SAP (HFA II, SITA Standard 24-2 program) was performed twice in the test eye prior to any experimental tests, or once if participants had undertaken one of these tests within the past six months as part of their routine clinical care; this ensured that all participants had adequate perimetric experience before undertaking experimental tests. False positive rates were < 15% for all participants.

Participants did not have any other ocular/systemic disease and/or medication known to affect visual performance (e.g. diabetes, thyroid disease, age-related macular degeneration, hydroxychloroquine medication); ocular health was confirmed by slit lamp biomicroscopy at each visit. One participant with glaucoma had previously undergone trabeculectomy surgery in the test eye seven years prior to the study; this eye had been considered stable by the hospital eye service since the surgery. Otherwise, no participants had undergone ocular surgery, with the exception of cataract removal. All participants had an IOP < 21 mmHg at all visits, measured with Goldmann Applanation Tonometry. Healthy participants did not have any first-degree relatives with glaucoma, and did not have any history of elevated IOP.

All participants had a best-corrected visual acuity of ≥ 6/9 (confirmed at all visits), in the absence of significant corneal or media opacities (≤ NO3, NC3, C3, and/or P3, Lens Opacities Classification System III; Chylack et al. 1993), with a spherical refractive error between +6.00 DS and -6.50 DS, and astigmatism < 3.50 DC in the test eye, as determined by a full refraction conducted before the commencement of any experimental tests. In those participants who had previously undergone cataract surgery, participants with pre-surgical refractive errors that did not meet these criteria were also excluded, if known. All experimental tests were conducted with natural pupils, and with participants wearing full refractive correction for a working distance of 30 cm.

One eye of each participant was tested. The test eye was selected as the eye that best met the inclusion/exclusion criteria, or was selected at random if both eyes were equally suitable. Of the participants with glaucoma, 15 right eyes and 15 left eyes were tested, and of the healthy participants, 11 right eyes and 9 left eyes were tested. Participants completed each of the four experimental tests on four separate visits (in randomised order) within a four-month period.

Ethical approval for the study was granted by the East of Scotland Research Ethics Committee (NHS Scotland). The research adhered to the tenets of the Declaration of Helsinki. Written, informed consent was obtained from all participants prior to inclusion.

4.2.2 Apparatus and set-up

Full details of the set-up and apparatus used are provided in chapter two. All stimuli were displayed on a gamma-corrected, 25" OLED display (Sony PVM-A250 Trimaster El, resolution 1920 x 1080 pixels, frame rate 60 Hz, refresh rate 120 Hz), driven by a ViSaGe MKII Stimulus Generator (Cambridge Research Systems, Rochester, UK). Experiments were programmed in MATLAB (version 2014b; The MathWorks, Inc., Natick, MA) using the CRS toolbox (version 1.27, Cambridge Research Systems, Rochester, UK), and adapted from code supplied by T. Redmond and P.J. Mulholland. A nominally uniform background luminance of 10 cd/m² was used. During all tests, participants maintained the correct viewing distance by placing their chin on a chin-

rest, and head against a head-rest, in a similar manner to conventional SAP (*Figure 4.2*). Participants wore a half-eye trial frame with full refractive correction incorporated using full aperture trial lenses, and the non-test eye occluded with a patch. During all tests, participants were instructed to fixate a central cross on the display screen, and respond to any stimulus they had detected in their visual field by pressing a button on a response pad (Cedrus RB-530; Cedrus, USA).
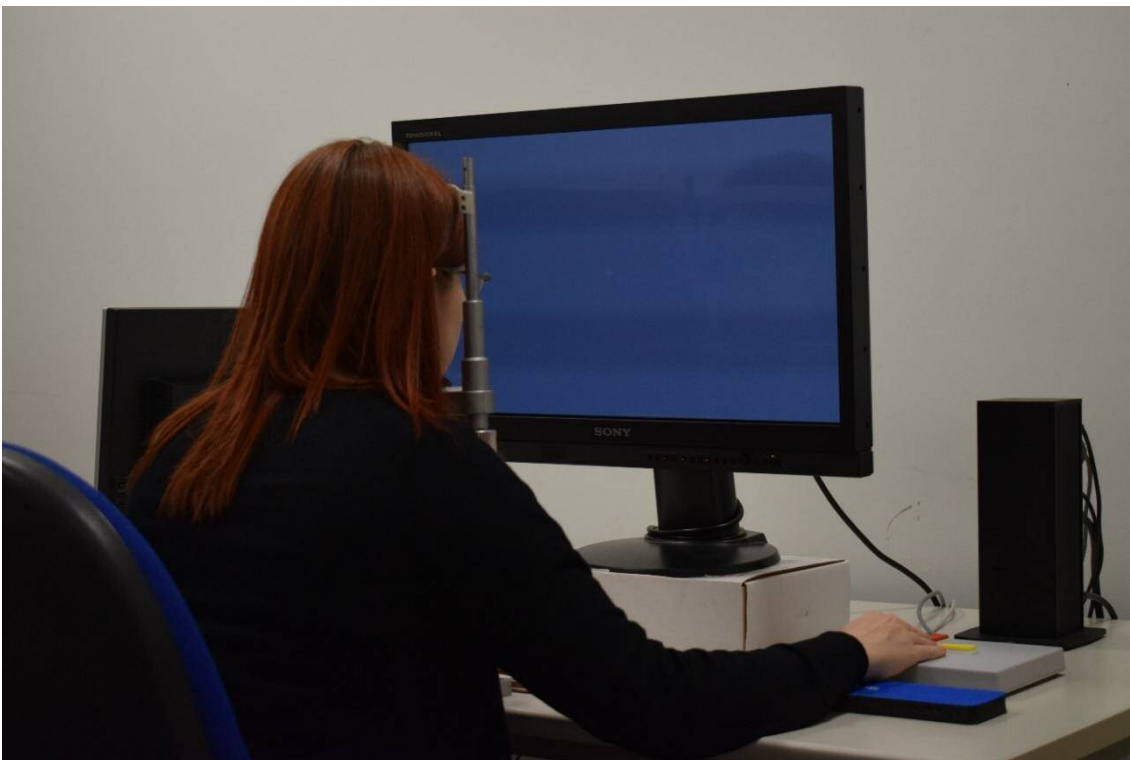
Figure 4.2 – Experimental set-up. Observers maintained the correct viewing distance by placing their chin on a chin-rest, and forehead against a head-rest. A half-eye trial frame was worn, with full refractive error corrected with full aperture trial lenses. The non-test eye was occluded with a patch. All tests were conducted with the room lights switched off.

In order to directly compare the performance of each stimulus form, all stimulus steps were converted to a common scale with identical units, according to *Equation 3.3*. Steps sizes were approximately equal, in terms of log energy, across stimulus forms, and with a common reference value.

Four visual field locations were tested, 9.9° from fixation along the 45°, 135°, 225°, and 315° meridians, as shown in *Figure 4.3*. The four different stimulus forms compared in this study, varied during experiments as described in section *4.2.3*.
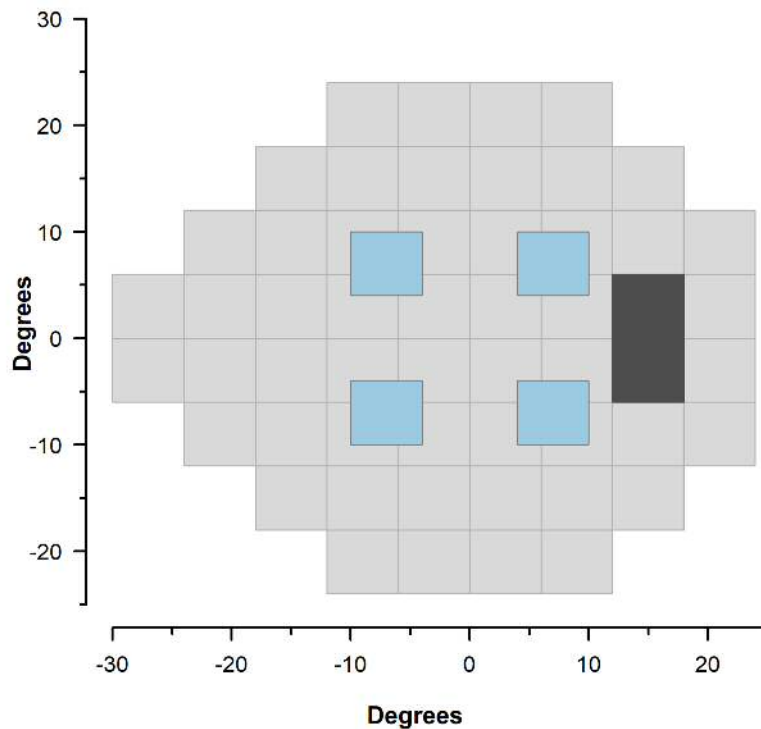
Figure 4.3 – The four test locations used in this experiment (shown in blue), in relation to a standard 24-2 test grid (shown in grey), for a right eye.

4.2.3   Stimuli

4.2.3.1 Contrast only (within Ricco's area, 'C_R')

The C_R stimulus was of a fixed area, within Ricco's area, varying in contrast. The area of this stimulus was determined from the study of Redmond et al. (2010a), in which Ricco's area was measured in those with early glaucoma, and healthy, age-similar participants, at four test locations 10° from fixation, along the 36°, 144°, 216°, and 324° meridians, similar to the test locations used here. Findings from this study are shown in *Figure 4.4*; Ricco's area values have been averaged for the superior and inferior hemifields, and for the purposes of this examination, both superior and inferior data are displayed on the same plot. The 0.1 percentile of Ricco's area values for healthy

participants is indicated by the lighter arrow; this corresponds to a stimulus area of -1.92 log deg^2 (0.01 deg^2). A stimulus of this area fell within Ricco's area for 99.9% of healthy participants, and 100% of glaucoma participants who took part in the study of Redmond et al. (2010a). As such, this was the stimulus area chosen for the C_R stimulus. To relate this to the standard Goldmann stimuli, this is slightly larger than a Goldmann I (-2.02 log deg^2), but smaller than a Goldmann II (-1.42 log deg^2).
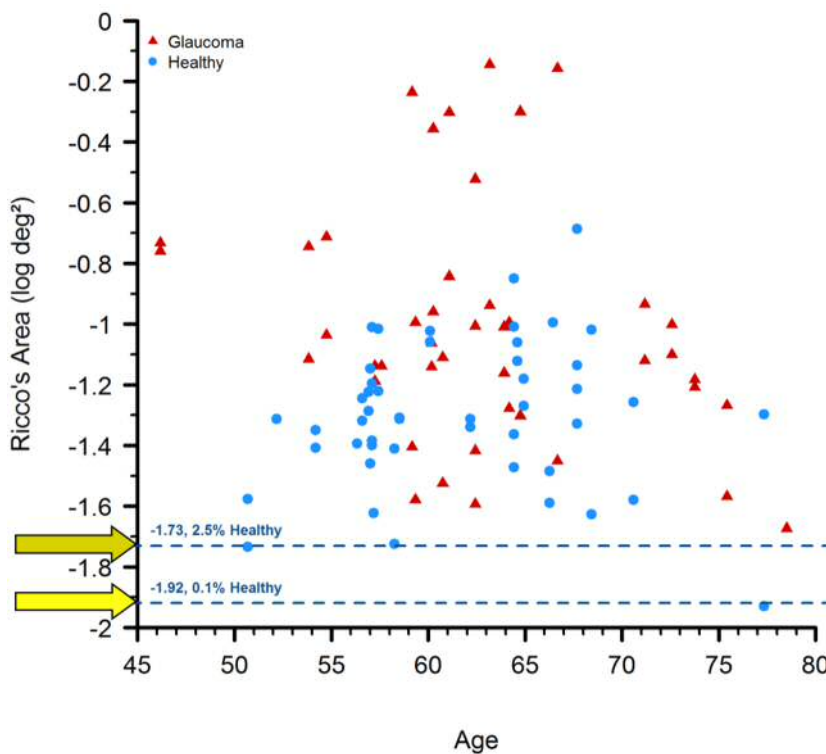
Figure 4.4 – Ricco's area measurements from the study of Redmond et al. (2010a), averaged across superior and inferior hemifields. For the purposes of this examination, both superior and inferior data is shown. The 0.1 and 2.5th percentiles for healthy participants are denoted by the dashed lines, and the lighter and darker arrows respectively.

Possible incremental stimulus contrast ranged from -1.66 log contrast (ΔI, 0.22 cd/m²) to 1.30 log contrast (ΔI, 198.85 cd/m²); this was the maximum available increment luminance on the Sony PVM-A250 Trimaster El OLED display, operating at 70% brightness and 80% contrast, as discussed in chapter two.

4.2.3.2 Area only ('A')

The A stimulus was of a fixed contrast, starting from within the area of complete spatial summation, varying in area. The contrast of this stimulus was determined such that it was equivalent to the threshold measurement for a stimulus of -1.73 log deg²; this is the 2.5[th] percentile of Ricco's area values for healthy participants (indicated by the darker arrow in *Figure 4.4*).

The smallest A stimulus was considered to be beyond the visibility of all observers. To ensure the contrast was set at an appropriate level to achieve this, the highest measured sensitivity (i.e. the lowest luminance) for a stimulus area of -1.73 log deg^2 was established from Redmond et al. (2010b). *Figure 4.5* shows a schematic of the average spatial summation curves for achromatic stimuli in the inferior and superior hemifields, for young, healthy observers aged 20-29 years (Redmond et al. 2010b). For the purposes of this examination, only the section of the curve indicating complete spatial summation is shown. Luminance threshold for a stimulus of -1.73 log deg^2 was established as 0.69 log ΔI/I (increment luminance of 4.84 cd/m^2) in the inferior hemifield, and 0.71 log ΔI/I (increment luminance of 5.12 cd/m²) in the superior hemifield. The luminance of the A stimulus was therefore set as the mean of these two values (0.70 log ΔI/I, increment luminance of 4.98 cd/m^2, log contrast ΔI -0.30).

Possible stimulus areas ranged from -2.52 log deg^2 (0.003 deg^2, i.e. the area of one pixel) to 2.16 log deg^2; this was the maximum area that could be presented without crossing horizontal or vertical midlines, or overlapping adjacent test locations.
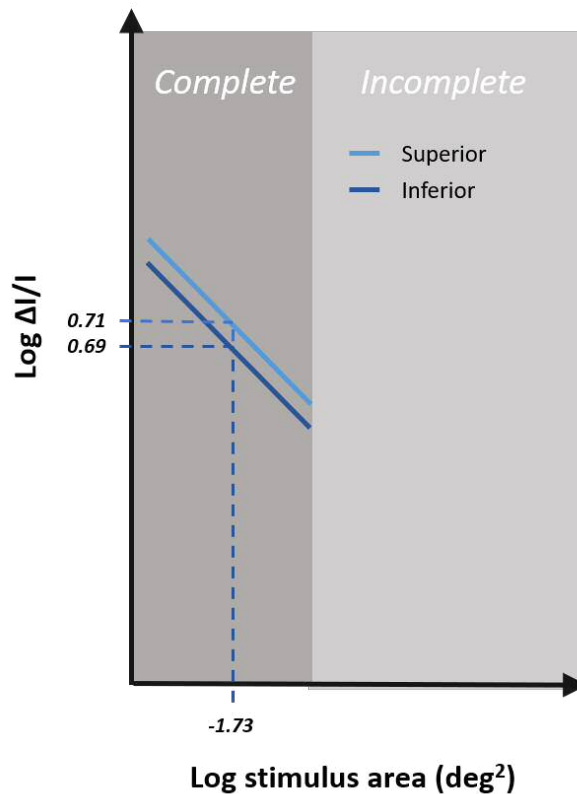
Figure 4.5 – Schematic showing average spatial summation curves (complete spatial summation only) for young, healthy observers (20-29 years), using achromatic stimuli. Findings are shown for both superior and inferior hemifields from the study of Redmond et al. (2010b). This was used to determine the lowest threshold for a stimulus of -1.73 log deg^2 (2.5th percentile for healthy participants from Figure 4.4). This diagram represents a schematic only, and is not meant to imply that Ricco's area measurements are equal in superior and inferior hemifields.

4.2.3.3 Area and contrast simultaneously ('AC')

The AC stimulus varied simultaneously and proportionally in both area and contrast, such that the slope of modulation was +1 in contrast/area space (as per *Figure 4.1*); as it has been reported that the most reliable data is obtained in kinetic perimetry when approaching the isopter perpendicularly (Lynn et al. 1991), it was speculated that approaching the spatial summation curve perpendicularly may reduce variability. The minimum stimulus had an area of -2.52 log deg^2 (0.003 deg^2), with a log contrast, ΔI, of -0.98 (increment luminance of 1.05 cd/m^2), and the maximum stimulus had an area of -0.11 log deg^2 (0.77 deg^2), with a log contrast, ΔI, of 1.28 (increment luminance of 188.60 cd/m^2). As this stimulus modulated simultaneously and proportionally in both area and contrast, the maximum area was limited by the luminance capabilities of the OLED display.

4.2.3.4 Contrast-only (Goldmann III-equivalent stimulus, 'GIII', reference stimulus)

As with the C_R stimulus, the GIII stimulus also varied in contrast, while maintaining a constant area; the area of the GIII stimulus was -0.95 log $\deg^2$ (0.11 $\deg^2$), similar to the stimulus area employed by commercial perimeters (Goldmann III). The same contrast scale was used as for the C_R stimulus.

For each stimulus form, a logarithmic 'look-up table' (LUT) was constructed of available stimulus parameters (luminance or area). In order to directly compare the performance of each stimulus form, all stimulus 'steps' were converted to a common scale with identical units, according to *Equation 3.3*. Stimulus duration was fixed at 0.2 seconds in all experiments. LUTs were constructed such that log stimulus energy in each step was approximately equal for the A, AC, and C_R stimuli; there were small differences between stimulus forms, particularly at the lower end of the energy scale, due to limitations in the stimulus area that could be displayed, owing to screen resolution.

In order to directly investigate the effect of stimulus configuration on disease signal, response variability and SNR, it was necessary to control, as much as possible, for any artefactual bias that could arise from the method used to determine these parameters. For example, it was necessary to control for a situation in which a test with one stimulus form could contain more supra-threshold presentations than one with another stimulus form, resulting in an artefactual steepening/flattening of the psychometric function. Thus, stimulus visibility was equated across all stimulus forms for each observer, as described in section *4.2.4*.

4.2.4 Psychophysical procedure

As in the experiment described in chapter three (section *3.2*), psychometric functions were measured at each location with each of the stimulus forms (separately) with a MOCS procedure. In an attempt to maximise efficiency, minimise slope bias, and equate stimulus visibility across all conditions, this experiment adopted a similar protocol to that of chapter three, whereby the psychometric function was densely sampled around the expected p(seen) = 0.5 region (50% seen, energy threshold), guided by the work of Hill (2001), with sufficiently supra- and sub- threshold stimuli

presented on the expected position of the asymptotes (*Figure 4.6*). To do this, FOS

curves were constructed using a three-stage approach, as described in sections *4.2.4.1*

to *4.2.4.3*; see *Figure 4.6* for an illustrated guide.
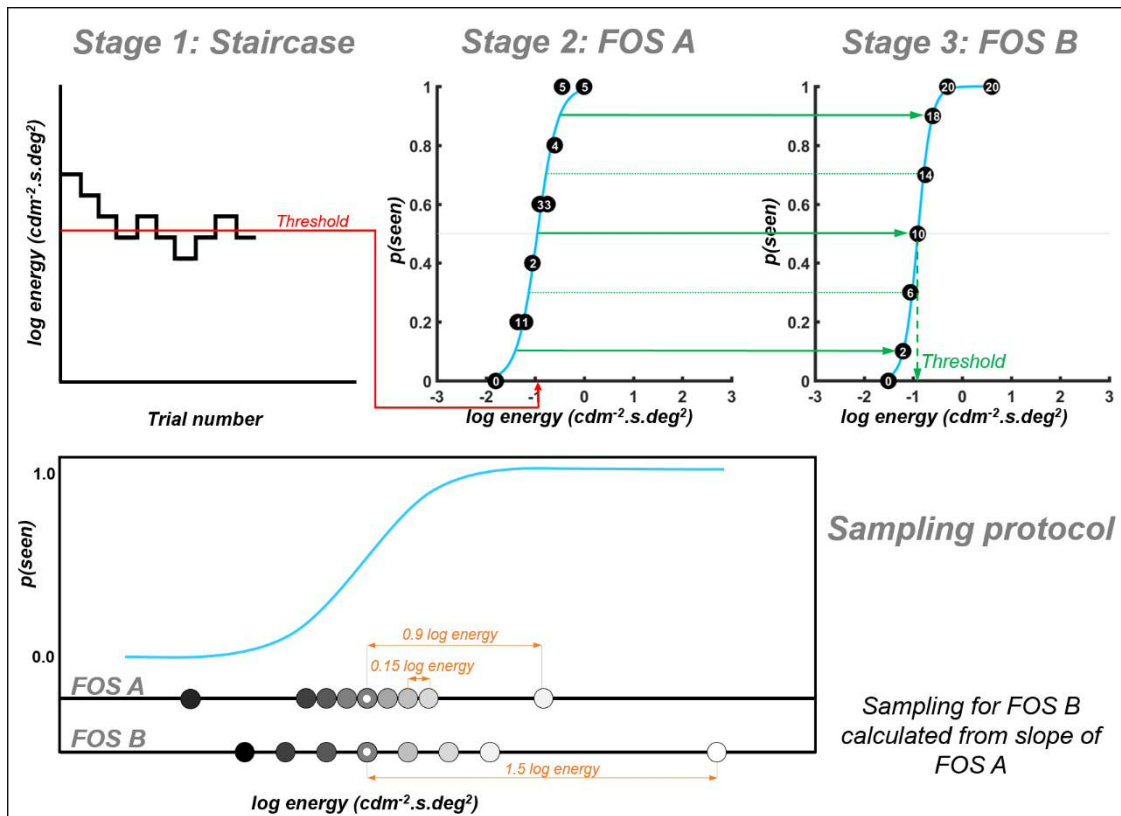
Figure 4.6 – Top: schematic of the three-stage process for finding threshold and response variability for each stimulus form. Stage one: short 1:1 staircase procedure. Stage two: short MOCS (five presentations per level), using the threshold from stage one to selected presented energy values. Stage three: standard MOCS (20 presentations per level), using FOS slope from stage two to inform the presented energy values (see sections 4.2.4.1 to 4.2.4.3 for a full description). Bottom: Illustration of the sampling protocol for FOS experiments (Rountree et al. 2018).

4.2.4.1 Stage One – Staircase procedure

To plan a sampling protocol for the MOCS (stages two and three, sections *4.2.4.2* and

4.2.4.3) it was necessary to perform a short 1:1 staircase procedure to determine an

approximate energy threshold for each of the four test locations (interleaved). The

staircase terminated after six reversals. Stimulus energy increased/decreased by an

average of 0.5 log energy following the first reversal, with proportionally smaller step

sizes following each subsequent reversal. Energy was modulated in 0.05 log unit steps

(the minimum possible step size) following the fourth and fifth reversals. The threshold at each location was taken as the mean of the final four reversals. The staircase procedure was performed twice, allowing participants the opportunity to become familiar with the stimulus form. Energy threshold values from the second test were then used in stage two.

4.2.4.2 Stage Two – Short MOCS phase

The purpose of this stage was to determine an approximate FOS curve position and slope, in order to optimise sampling of the curve in stage three. This stage consisted of a short MOCS procedure, using nine energy levels, each presented five times at each of the four test locations (180 presentations in total). The nine energy levels were the energy threshold from stage one, three above and three below this initial energy threshold value, each separated by 0.15 log energy, and two further values, 0.9 log energy above and below the initial energy threshold level.

Presentations were randomised in terms of energy level and test location. A rest break was taken halfway through the test (after 90 presentations). A FOS curve was constructed from the results for each of the four test locations, and fitted with a psychometric function. Energy levels at p(seen) = 0.1, 0.3, 0.5, 0.7, and 0.9 were estimated from the curve and used to sample the psychometric function in stage three.

4.2.4.3 Stage Three – Standard MOCS phase

In this stage, participants were presented with 20 repetitions of eight energy levels at each of the four test locations (640 presentations in total). At each location, five of the energy levels were determined from the FOS curve for the same location in stage two (section *4.2.4.2*, values for p(seen) = 0.1, 0.3, 0.5, 0.7 and 0.9). If one or more of the energy levels could not be established from the short MOCS phase, due to incompletion of the FOS curve, an interval of 0.25 log energy was used between these levels. Three additional energy levels were presented; two levels were p(seen, 0.5) ± 2 SD from the psychometric function (according to *Equation 3.2*) in stage two, rounded to the nearest available energy interval, and one additional level high above (p(seen, 0.5) + 1.5 log energy), to ensure that a greater number of stimuli were supra-threshold

than sub-threshold and thus aid observer attention. The energy levels at all test locations were randomly presented. A rest break was taken at every quarter (after 160 presentations). As discussed in chapter three, the resulting FOS data were fitted with a logistic psychometric function, with guess and lapse rates allowed to vary between 0 and 0.1 (0-10%). The energy threshold was established as the energy value at p(seen) = 0.5. Response variability was taken as the SD of the psychometric function (according to *Equation 3.2*). These values were used in subsequent analyses of signal, noise, and SNR.

Participants completed tests for all four stimulus forms; the order in which tests were undertaken was randomised for each participant. Participants could complete tests for up to two stimulus forms in any one day. If tests for more than one stimulus form were completed in one day, participants were given a rest break of 30 minutes from conclusion of tests with the first stimulus form, to commencement of tests with the second stimulus form. In addition to the scheduled rest breaks, participants could request additional breaks as required.

4.2.5 Statistical analysis

Fitting of psychometric functions, and analysis of FOS data were performed in MATLAB (version R2015b; The MathWorks Inc., Natick, MA, USA), using the Palamedes toolbox (Prins and Kingdom 2009). Analyses described from this point were conducted on those FOS data collected in stage three. Statistical analyses were performed with the freely available, open source statistical environment R (R Development Core Team, 2017), and SPSS (IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0, Armonk, NY: IBM Corp). To ensure appropriate comparisons between quadrants, all data were converted to that for a right eye.

4.2.5.1 Total deviation (TD)

To examine differences in disease signal between stimulus forms, energy thresholds for healthy participants were pooled across the four test locations, plotted against age for each of the four stimulus forms, and fitted with a mixed model linear regression. TD was then calculated for each test location in participants with glaucoma as the

difference between measured threshold and the expected threshold for that of an age-matched normal, estimated from the linear regression model for that stimulus form.

TD values for the GIII (Goldmann III-equivalent) stimulus were pooled across the four test locations, and divided into three TD strata: lower (between the 99th and 66th percentiles, equivalent to a localised perimetric sensitivity of > 28.4 dB with HFA II), middle (between the 66th and 33rd percentiles, equivalent to a localised perimetric sensitivity between 24.6 and 28.4 dB), and upper (within the 33rd percentile, equivalent to a localised perimetric sensitivity of < 24.6 dB). TD values for A, AC and C_R stimuli were plotted against those measured with the GIII stimulus, and the residuals for each stimulus from a line of equation $x = y$, (i.e. the GIII stimulus plotted against itself) were examined, to determine whether TD values were generally higher or lower than those with the GIII stimulus.

4.2.5.2 Response variability

To test the hypothesis that a stimulus varying in area has lower response variability compared to conventional stimuli, response variability was compared between all stimuli at each test location in the lower disease stratum with a Friedman test, with post hoc Wilcoxon signed-rank tests. In addition, to determine the association between response variability and disease severity for each stimulus form, a total least squares linear regression was performed on these data at each test location. As response variability was determined as the SD of the psychometric function, a response variability of zero represented the ideal observer (i.e. a purely vertical slope), with larger values representing a greater response variability. Therefore, in the total least squares analysis, steeper regression slopes indicate more marked dependence of response variability on TD, while a regression slope of zero indicates that response variability is largely independent of TD.

4.2.5.3 Signal/noise ratio (SNR)

As neither disease signal, nor response variability alone can fully inform the utility of one stimulus over another, SNR (TD/response variability) was compared between stimulus forms, and across the three disease strata. A linear mixed effects model analysis of the relationship between SNR and stimulus form was performed on SNR

data pooled from all four test locations, using the lme4 package (Bates et al. 2015). Stimulus form and stratum (without an interaction term) were entered as fixed effects. Intercepts for subjects and test locations, as well as by-subject random slopes for the effect of stimulus forms, were entered as random effects. There were no obvious deviations from normality, nor heteroskedasticity. Likelihood ratio tests of the model including the effect in question (SNR), against the same model excluding the effect, were used to determine one-tailed p-values.

In all statistical analyses, a Holm-Bonferroni post hoc correction was applied where there were multiple tests of the same hypothesis. All p-values quoted here have been post hoc corrected.

4.2.6   Fatigue effect and repeatability

Of the healthy participants, three completed additional tests to investigate fatigue effect and repeatability. These participants were aged 72.3, 74.1, and 78.1 years, with MD measurements of +1.96, -0.36, and +0.26 dB with the SITA Standard 24-2 strategy (HFA II). Two right eyes and one left eye were tested.

These participants completed the tests for each of the four stimulus forms as described above. In addition, they returned on four further days, an average of nine months after the original tests; refractive error was re-determined, and participants again completed a SITA Standard 24-2 strategy (HFA II). Visual acuity, IOP with Goldmann Applanation Tonometry, and slit lamp biomicroscopy were conducted at each visit, thus ensuring that no change had taken place in the interim. On each of the four days, participants completed tests for the same stimulus form twice, following the same three-stage psychophysical procedure as described in section *4.2.4*, with a rest break of 30 minutes from conclusion of the first test to commencement of the second test.

To evaluate repeatability of the four stimulus forms, threshold and response variability values were pooled across the four test locations, and were analysed across the three tests. A linear mixed effects model analysis was performed of the relationship between threshold and test number, and response variability and test number, with two-tailed p-values determined from a likelihood ratio test.

4.3 Results

As a clinical indicator of the severity of local damage tested in this experiment, raw dB and TD values were predicted from the final preliminary SAP test (SITA Standard 24-2, HFA II) using bi-linear interpolation in MATLAB, as the four test locations are not part of a standard 24-2 test grid (*Figure 4.3*). An example is shown for one of the participants with glaucoma in *Figure 4.7*; perimetric sensitivity values for the 52 HFA II test locations are displayed in *Figure 4.7.A* (minus the two adjacent to the blind spot), and the interpolated perimetric sensitivity values for the four test locations are indicated in bold italics. The same is shown for TD values in *Figure 4.7.B*. Histograms in *Figure 4.8.A* and *Figure 4.8.B* show the interpolated perimetric sensitivity values (dB) and TD values (TD$_{SAP}$) respectively for the 30 participants with glaucoma, and *Figure 4.8.C* and *Figure 4.8.D* show the same values for the 20 healthy participants.

Examples of the psychometric functions obtained are given in Appendix A. Mean guess rate was 0.01 (1%) for both healthy and glaucoma participants, and mean lapse rate was 0.02 (2%) and 0.04 (4%) for healthy and glaucoma participants respectively.
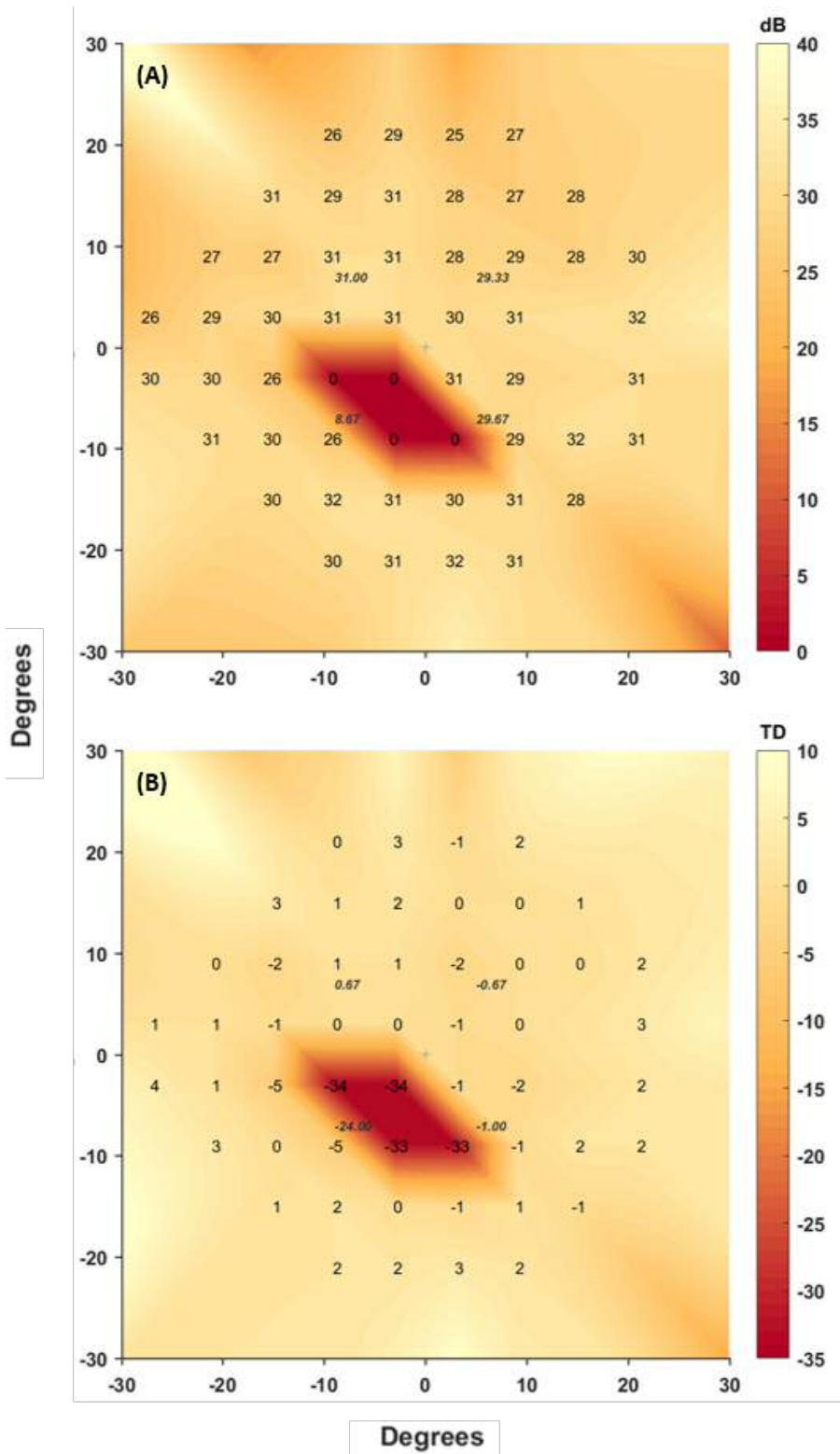
Figure 4.7 – Values from HFA II SITA Standard 24-2 for one participant with glaucoma, showing (A) raw dB values, and (B) TD values. Interpolated values for the four test locations are indicated in bold italics.
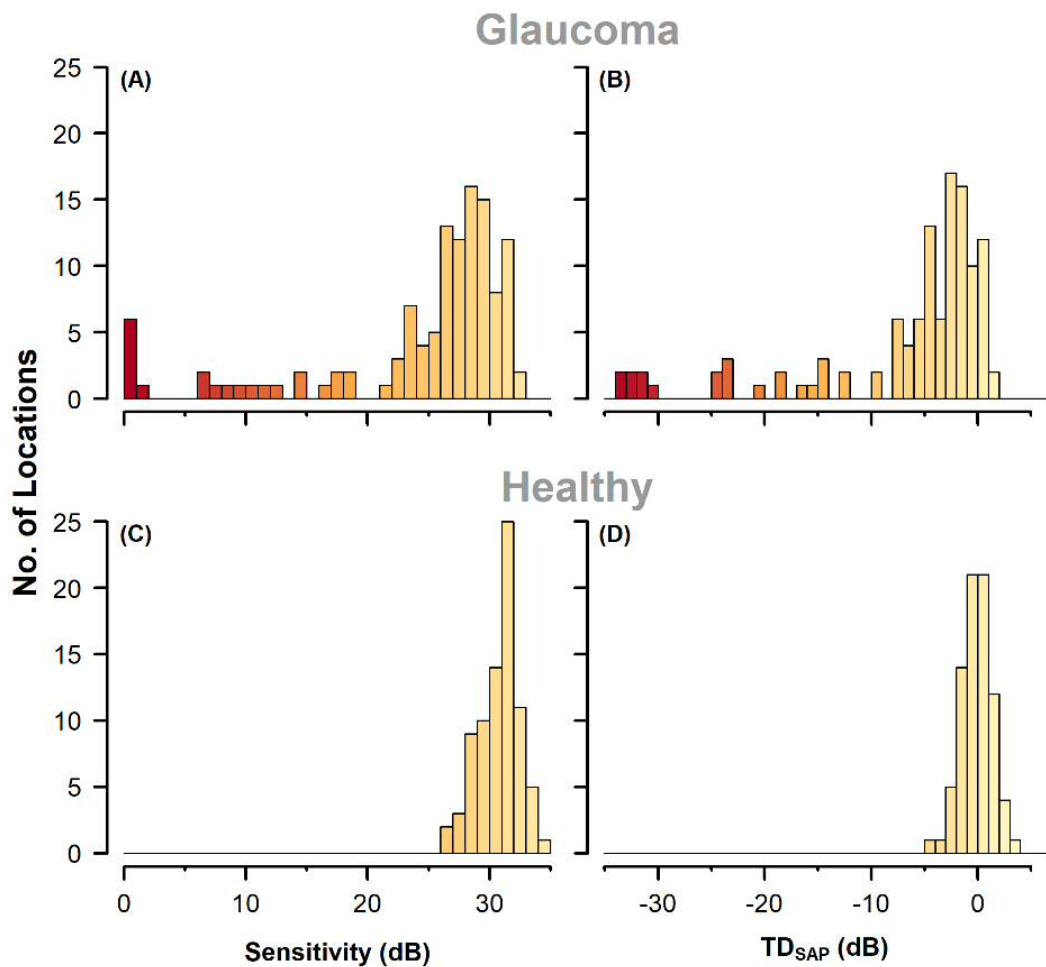
Figure 4.8 – Perimetric sensitivity (SITA Standard 24-2, HFA II), interpolated for the four test locations used in this experiment (Rountree et al. 2018). (A) Raw dB values for participants with glaucoma. (B) TD values for participants with glaucoma. (C) Raw dB values for healthy participants. (D) TD values for healthy participants.

4.3.1 Total Deviation

Energy threshold, p(seen, 0.5), was plotted against age for healthy participants at each of the four test locations, per stimulus form. Each stimulus form was fitted with an ordinary least squares (OLS) linear regression, as shown in *Figure 4.9*. The interpolated perimetric sensitivity values from the HFA II SITA Standard 24-2 (as per *Figure 4.7*) for each healthy participant were converted to threshold log energy in accordance with *Equation 3.3*, to be consistent with the four stimulus forms used in the experimental tests; these values are also plotted in *Figure 4.9*. Threshold differences with age were observed to be similar at all four test locations, with similar linear regression slopes. Therefore, threshold values were pooled for the four test locations and plotted against

age (*Figure 4.10*). As thresholds were not normally distributed with all four stimulus forms at each test location (Shapiro-Wilk test), a Friedman analysis, with post hoc Wilcoxon signed-rank tests, was conducted separately for each stimulus form to compare threshold differences between the four test locations. This did determine some statistically significant differences in energy threshold between some locations with the A, GIII and HFA stimuli. To account for this, and to account for the reduced independence between data points, a linear mixed effects model was fitted to the data. TD was then calculated for each glaucoma test location as the difference between measured energy threshold and that of an age-matched normal, estimated from the mixed model linear regression for that stimulus form.
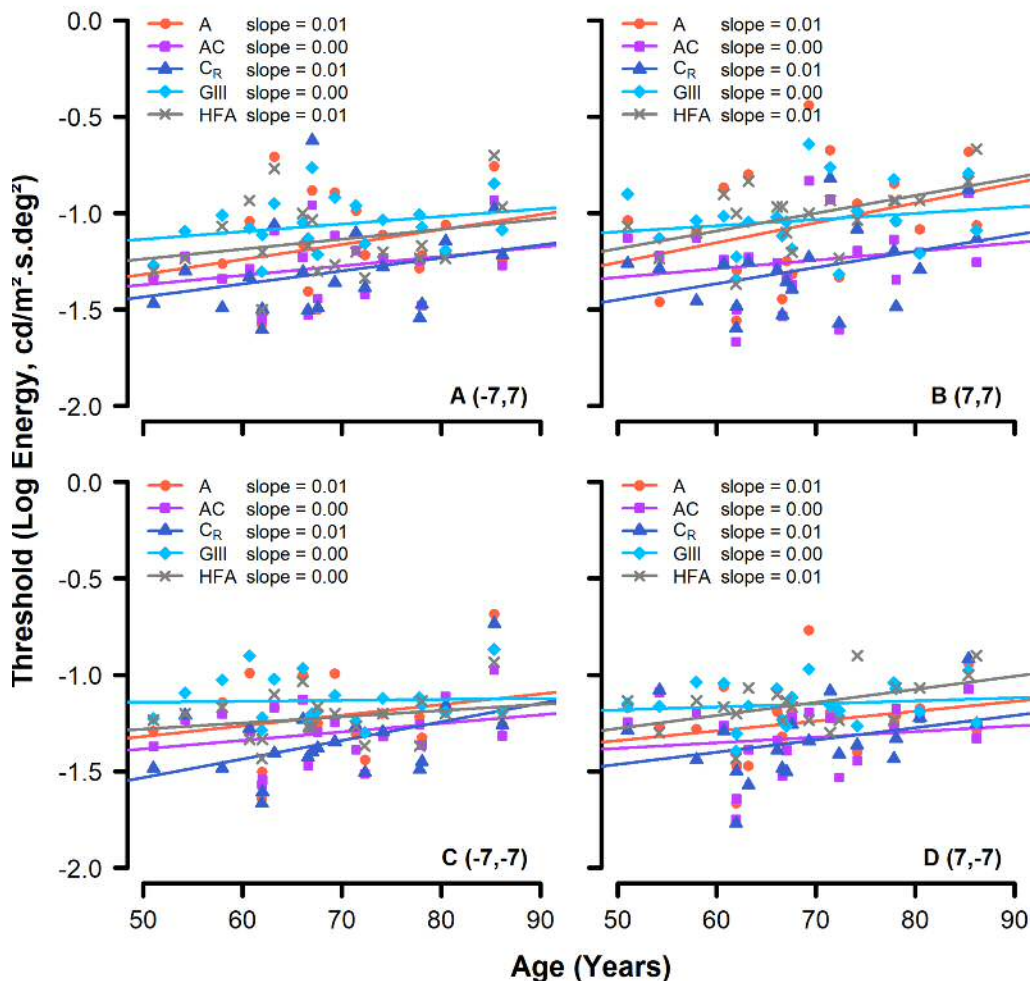
Figure 4.9 – Energy threshold plotted against age for healthy participants at each of the four test locations. Each stimulus form was fitted with an ordinary least squares linear regression. Slopes for this regression are shown for each stimulus form.
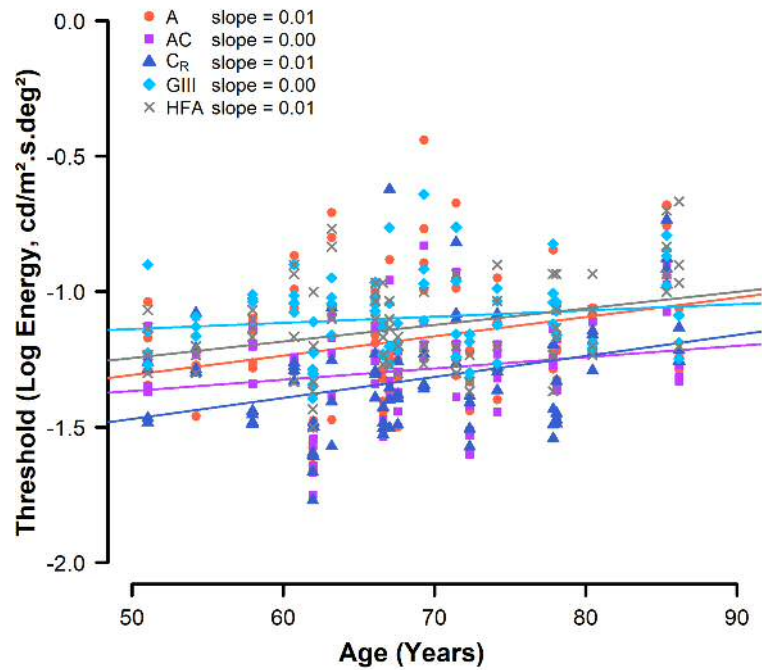
Figure 4.10 – Energy threshold, pooled for the four test locations and plotted against age for healthy participants. Each stimulus form is fitted with a mixed model linear regression. Slopes from this regression are shown for each stimulus form.

Of the 120 test locations across the glaucoma cohort (30 participants, four test locations), energy threshold, and therefore TD, could not be established at two locations for the A stimulus, 15 for the AC stimulus, 42 for the C_R stimulus, and 21 for the GIII stimulus. This was due to incomplete FOS curves, in that a p(seen) = 0.5 value could not be reliably determined; differences between stimulus forms reflect the differing dynamic ranges with the apparatus used. There were no such incomplete FOS curves with any of the four stimulus forms in the healthy cohort. To compare stimulus forms directly, independent of dynamic range, a separate analysis was conducted on only those test locations whereby TD could be established with all four stimulus forms ('matched' data), in addition to analysing all available data ('complete' data).

Figure 4.11 shows TD values for complete (*Figure 4.11.A*) and matched (*Figure 4.11.B*) data for the four stimulus forms in participants with glaucoma. TD for each stimulus form has been plotted against TD for the reference (GIII) stimulus. The light-blue line indicates TD for the GIII stimulus (i.e. plotted against itself, $x = y$), and as such is used

as a reference line. Data points above this reference line indicate a greater TD for that stimulus form than for the GIII stimulus, and those below the reference line indicate a lower TD. 'L', 'M' and 'U' denote the lower, middle and upper strata, according to the TD for the GIII stimulus (as described in section *4.2.5.1*). A negative value for TD denotes a lower energy threshold than that of the age-matched normal threshold for that stimulus form. 'Unfilled' data points denote those in which TD could not be established with the GIII stimulus, but could be established with other stimulus forms. These stimulus TD values were plotted instead against TD calculated from perimetric sensitivity measured with the SITA Standard 24-2 program on the HFA II, as per *Figure 4.10*. The unfilled data points are presented for illustration purposes only, and were not used in further analysis.

The residuals of the data points in *Figure 4.11* (i.e. the difference between TD for each of the A, AC, and C_R stimuli, and TD with GIII) were calculated for both complete and matched data (excluding unfilled data points). These were then averaged, to indicate whether there was an overall larger, or smaller, disease signal with each of the three test stimuli, compared to that with the GIII, and which stimulus form gave the greatest increase. These values are shown in *Table 4.1*; average residuals for A, AC and C_R were all positive, indicating that, overall, TD and therefore disease signal with all three test stimulus forms was higher than that with the GIII. This was true for both complete and matched data, with the A stimulus showing the largest overall disease signal (0.14, matched data).
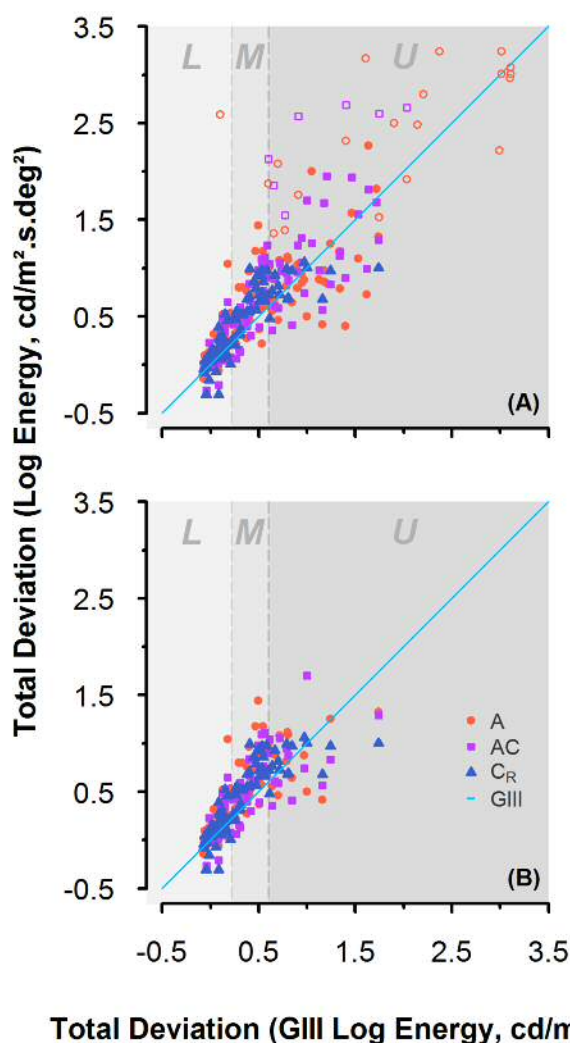
Figure 4.11 – TD values for each stimulus form, plotted against TD for the GIII stimulus (Rountree et al. 2018), for (A) complete, and (B) matched data, pooled across the four test locations. Light-blue line: TD with the GIII (reference) stimulus. L, M, U: Lower, middle, and upper strata representing three levels of disease severity studied here. Unfilled data points: interpolated SAP sensitivity converted to TD, where TD could not be measured with the GIII on the experimental apparatus.

	Average residuals (TD difference from GIII)
	Complete
A	0.08
AC	0.09
C$_R$	0.09
	Matched
A	0.14
AC	0.08
C$_R$	0.09

Table 4.1 – Mean residuals (TD difference from reference line, GIII in Figure 4.11) for each of the three stimulus forms (Rountree et al. 2018). As all numbers are positive, this indicates an overall increase in TD for each of the three stimulus forms, for both complete and matched data. Unfilled data points in Figure 4.11 were not included in this calculation.

4.3.2   Response variability

Of the 120 test locations across the glaucoma cohort, SD, and therefore response variability, could not be established at three for the A stimulus, 17 for the AC stimulus, 49 for the C_R stimulus, and 24 for the GIII stimulus. As with threshold, this was due to incomplete FOS curves, such that SD could not be reliably determined at these locations.

Figure 4.12 shows the response variability for each stimulus form in the lower disease stratum only, plotted against TD for that stimulus form at each of the four test locations. Complete and matched data were the same in this stratum, i.e. a response variability value was achieved with all four stimulus forms, at all test locations. As measurements of response variability in the lower disease stratum were not normally distributed with all four stimulus forms at each test location (Shapiro-Wilk test), a Friedman analysis, with post hoc Wilcoxon signed-rank tests, was performed. The response variability was found to be statistically significantly higher with the A stimulus compared with the GIII stimulus at the [-7,-7] location only (p = 0.048). No other statistically significant differences were found between any other stimulus forms at any locations (p > 0.05 for all comparisons).
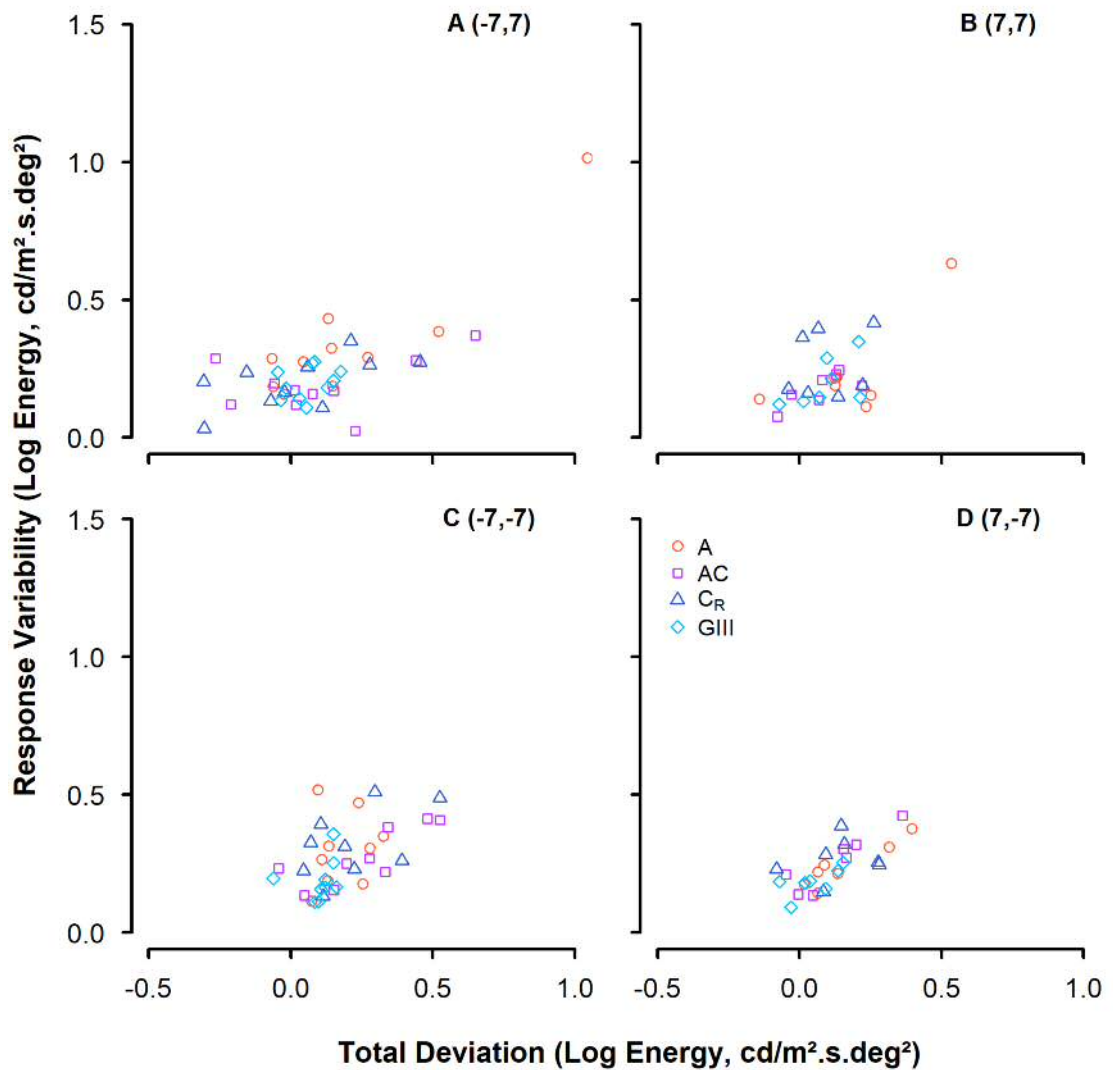
Figure 4.12 – Response variability for each stimulus form, plotted against its own TD for that stimulus form, for the lower disease stratum at the four test locations.

Figure 4.13 and *Figure 4.14* show the response variability at all levels of disease severity at the four test locations, for complete and matched data respectively. Response variability for each stimulus form has again been plotted against TD for that stimulus, to demonstrate how response variability is affected by disease severity. A total least squares linear regression model was fitted to the data, as TD (on the x-axis) was not fixed by the study design; slope values from this regression model are shown for each stimulus form.
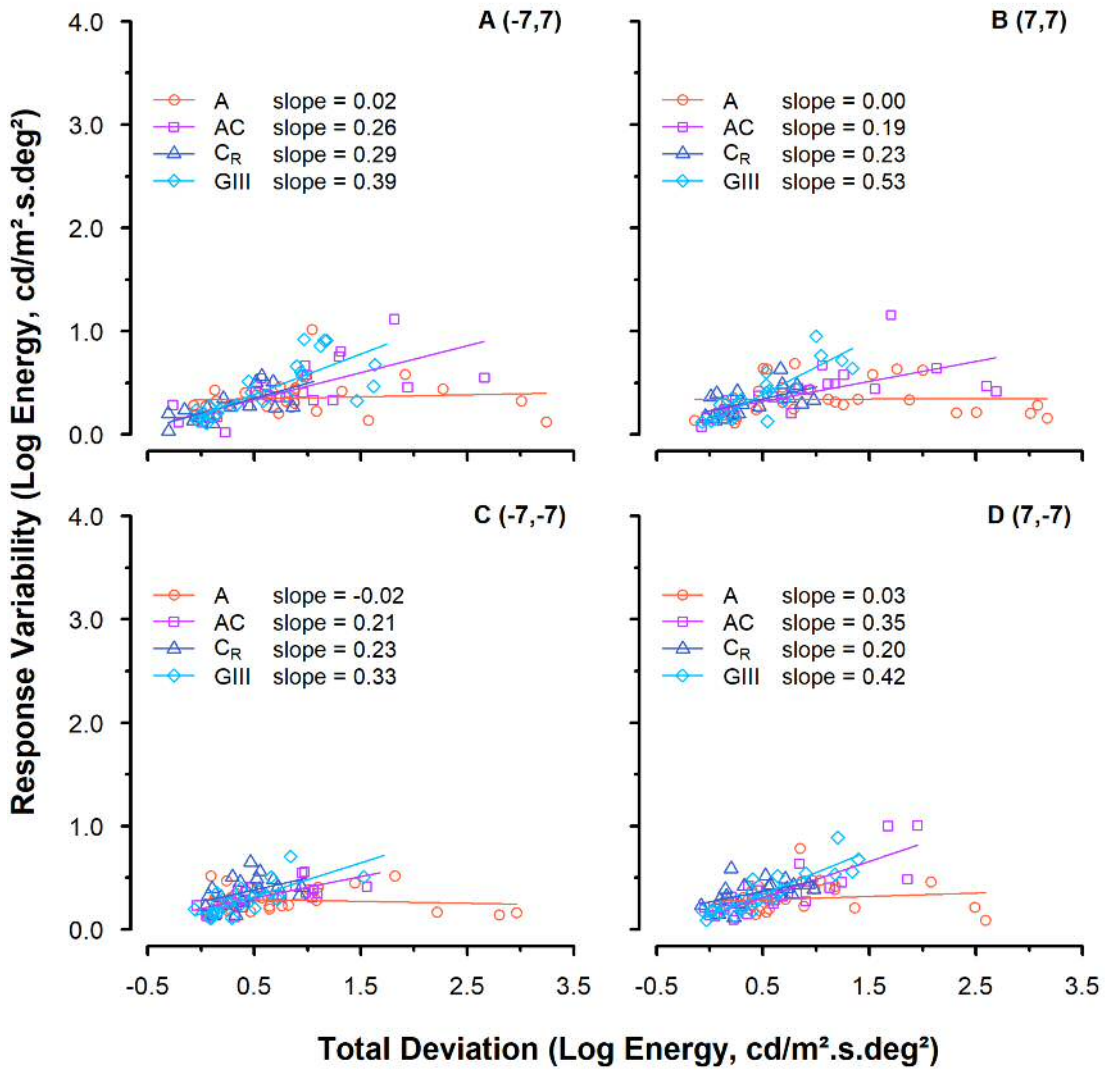
Figure 4.13 – Response variability for each stimulus form, plotted against its own TD for that stimulus form (Rountree et al. 2018). Data are shown for each of the four test locations (complete data) and are fitted with a total least squares linear regression model, with slope values displayed.
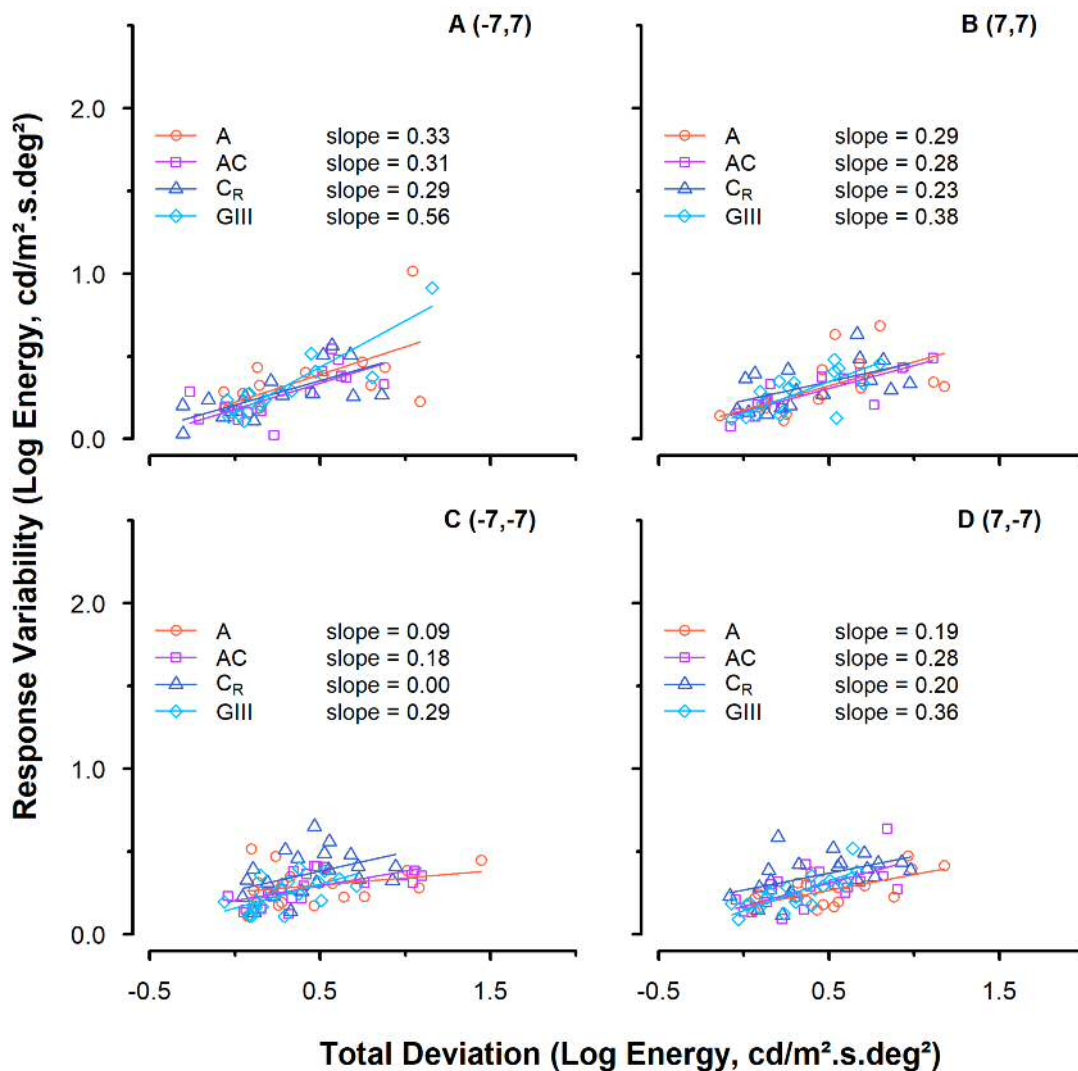
Figure 4.14 – Response variability for each stimulus form, plotted against its own TD for that stimulus form. Data are shown for each of the four test locations (matched data) and are fitted with a total least squares linear regression model, with slope values displayed. The y-axis is scaled up in comparison to Figure 4.13 for ease of data visualisation.

For both complete and matched data, total least squares regression slopes were steepest with the GIII stimulus at all four test locations, indicating that response variability was most dependent on depth of defect with this stimulus. For complete data, the shallowest slopes (least dependence on depth of defect) were found with the A stimulus at all locations, followed by AC, then C_R, in three out of the four locations. For matched data, the shallowest slopes were found with the C_R stimulus in three out of the four locations, and with the A stimulus at the other test location.

4.3.3   SNR

Figure 4.15 shows SNR for the four stimulus forms, pooled across all four test locations, and separated into the same three disease strata according to TD for the GIII stimulus (detailed in section *4.2.5.1*). SNR is shown for complete (*Figure 4.15.A*) and matched (*Figure 4.15.B*) data.

One-tailed p-values from the mixed model regression analysis (conducted for matched data only) are displayed in *Table 4.2*. Overall, when all three disease strata were considered together, both the A and AC stimuli had a statistically significantly higher SNR when compared with the GIII stimulus, by 0.66 ± 0.15 standard error (SE, $p < 0.01$), and by 0.25 ± 0.09 SE ($p = 0.008$) respectively. Overall, the C_R stimulus had a higher SNR than that for the GIII stimulus, by 0.11 ± 0.10 SE, but this was not statistically significant ($p = 0.27$). SNR for the A stimulus was higher than that for the GIII stimulus in each stratum ($p = 0.17$, 0.001, and 0.02 in the lower, middle, and upper strata respectively). SNR for the AC stimulus was consistently higher than that for the GIII stimulus in all three strata, though not by a statistically significant amount ($p = 0.5$, 0.06, and 0.07, in the lower, middle, and upper strata respectively). SNR for the C_R stimulus was higher than that for the GIII in all three strata, but this difference was not statistically significant ($p = 0.5$, 0.42, and 0.30 in the lower, middle, and upper strata respectively).
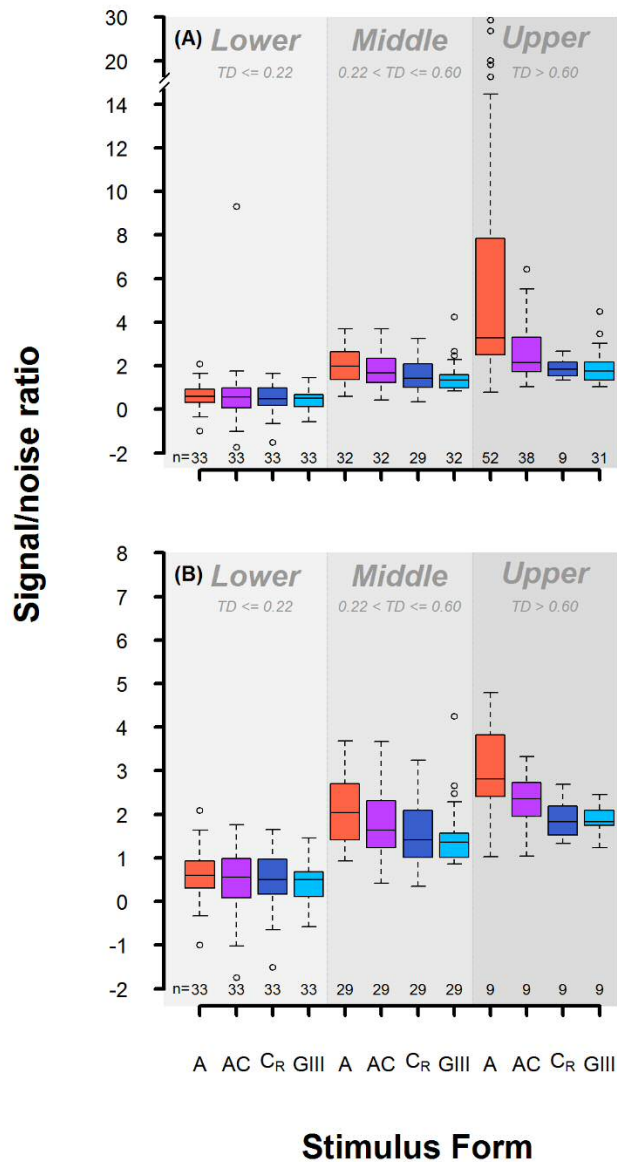
Figure 4.15 – SNR for each stimulus form, for three strata of disease severity, lower, middle and upper percentiles according to TD with the GIII stimulus (Rountree et al. 2018). Same strata as in Figure 4.10. Boundaries between strata are indicated in both log energy, and HFA II-equivalent sensitivity), for (A) complete and (B) matched data, pooled across all four test locations. Sample size (n) is given below each box. The y-axis in (B) is scaled up for ease of data visualisation.

Post hoc p-values (all strata)			
	A	**AC**	**C_R**	**GIII**
A		0.02*	0.004*	< 0.001*
AC			0.27	0.008*
C_R				0.27
GIII				

*Table 4.2 – Holm-Bonferroni corrected one-tailed p-values for overall differences in SNR between stimulus forms across the three disease strata, with mixed model regression analysis (Rountree et al. 2018). *Statistically significant at the p < 0.05 level.*

Figure 4.16 illustrates the difference in SNR between the test stimuli and the GIII stimulus, as a function of defect depth (TD for the GIII stimulus). It can be seen in *Figure 4.16.A* that in each stratum, the majority of data points lie above the line of equality, illustrating a greater SNR for the A than for the GIII stimulus at all stages of disease studied here. Effect sizes (difference in SNR from that of GIII, reported by the linear mixed effects analysis) for each of the three disease strata, are given in *Figure 4.16*, with SE in brackets. A systematically greater effect size can be seen between lower and upper strata for the A stimulus, while the differences in effect size across strata for the AC and C_R stimuli are more consistent.

Figure 4.16 shows two distinct outliers in the lower stratum, one for the AC stimulus (*Figure 4.16.B*), and one for the C_R stimulus (*Figure 4.16.C*), i.e. those falling outside the split y-axes. As these outliers were found to unduly influence the results of the linear mixed effects analysis, they were excluded from all analyses detailed here.

Given that the lower stratum likely included test locations that were completely unaffected by disease, as well as locations that were in the early stages of disease, the lower stratum was subdivided again to further examine differences between stimuli. Data from the lower stratum in *Figure 4.15*, are isolated in *Figure 4.17.A*. Complete and matched data were the same in this stratum, i.e. an SNR value was achieved with all four stimulus forms, at all test locations. From this dataset, the lower 10%, according to TD values for the GIII stimulus, were removed and the remaining 90% replotted and reanalysed with the linear mixed effects model. Then the lower 20%, according to TD values for the GIII stimulus, were removed and the remaining 80% replotted and reanalysed with the linear mixed effects model. This was conducted for every 10% of the data, the 20% intervals of which are shown in *Figure 4.17.B-E*. Although the difference between the A and GIII stimulus forms does appear to become more apparent with each subdivision, no statistically significant differences were found with the linear mixed effects analysis ($p > 0.05$ in all subdivisions).
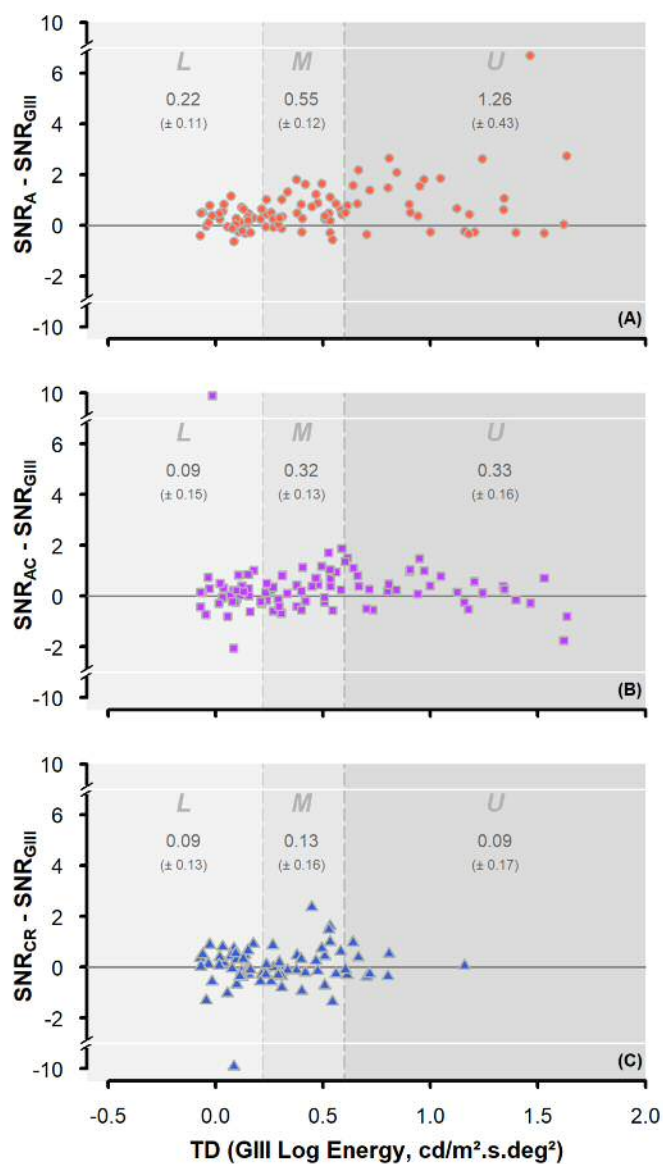
Figure 4.16 – SNR differences between (A) A and GIII, (B) AC and GIII, and (C) C_R and GIII, plotted against TD for GIII (Rountree et al. 2018). The three disease strata are indicated. Effect sizes from the linear mixed effects analysis are indicated, with SE in brackets.
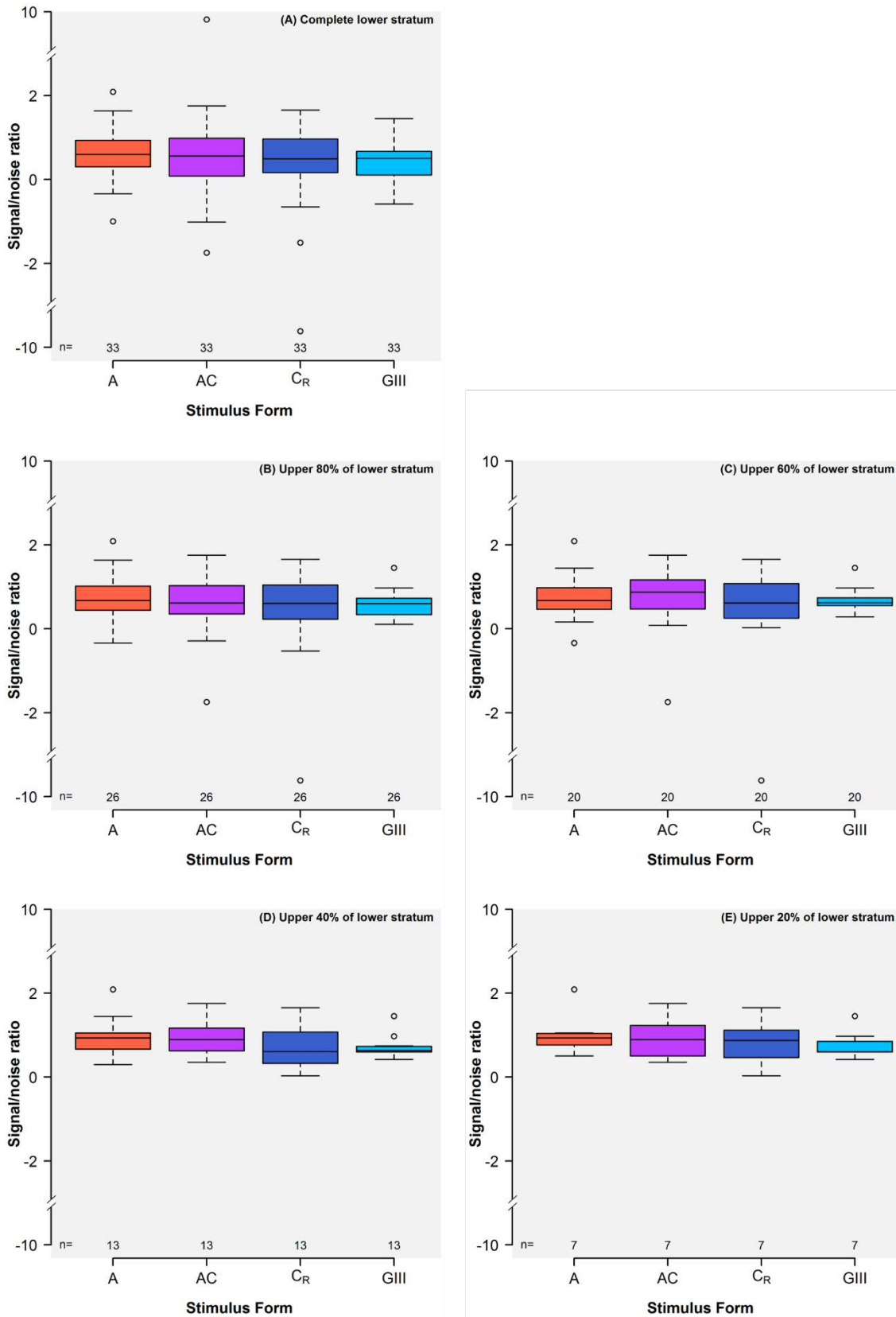
Figure 4.17 – (A) SNR for the lower stratum. (B)-(E) subdivisions of the SNR for the lower stratum, in which the lower (B) 20%, (C) 40%, (D) 60% and (E) 80% have been removed.

4.3.4   Fatigue effect and repeatability

Figure 4.18 shows box-and-whisker plots for energy threshold (*Figure 4.18.A*) and response variability (*Figure 4.18.B*) for each of the three tests conducted for the four stimulus forms; data from the four test locations have been pooled. Test one was conducted as part of the larger experiment, the results of which are included in sections *4.3.1* to *4.3.3*. Tests two and three were conducted on a separate day, with a rest break of 30 minutes between conclusion of test two, and commencement of test three.
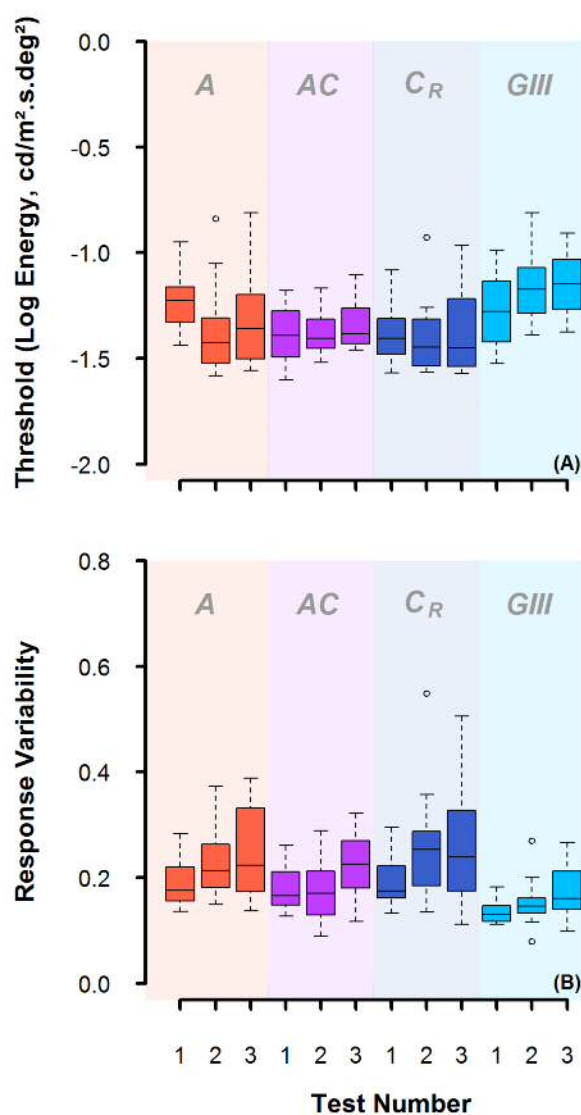
Figure 4.18 – (A) energy threshold and (B) response variability for three tests, for each stimulus form, pooled across the four test locations. Test one was conducted as part of the larger experiment, and tests two and three were conducted on a separate day, with a 30 minutes rest break between.

Energy thresholds appear slightly higher in the third test compared to the second test for all stimulus forms, i.e. the second of the two tests conducted on the same day. Differences between the first test and the second test are less consistent, with a higher threshold on the second day with some stimulus forms (A and GIII) and a lower threshold on the second day with others (AC and C_R). A linear mixed effects analysis was performed across the three tests, conducted individually for each stimulus form. No statistically significant differences were found between the three tests for any of the four stimulus forms (all p > 0.05), as shown in *Table 4.3*.

Response variability appears to be higher with all stimulus forms on the second day (test two) compared to the first day (test one). There is less consistency between tests two and three, i.e. the two tests conducted on the same day, with some stimulus forms showing a lower response variability in test three compared with test two (C_R), and others showing a higher response variability (A, AC, GIII). As for energy threshold, a linear mixed effects analysis was performed across the three tests, conducted individually for each stimulus form. No statistically significant differences were found between the three tests for any of the four stimulus forms (all p > 0.05), as shown in *Table 4.3*.

Stimulus Form	Threshold	Response variability
A	0.14	0.25
AC	0.42	0.09
C_R	0.88	0.34
GIII	0.72	0.08

Table 4.3 – T wo-tailed p-values from the linear mixed effects analysis of the relationship between threshold/response variability and test number. This was conducted individually for each of the four stimulus forms.

4.4   Discussion

In order to establish superior utility of a particular stimulus form over another for a given clinical purpose, a greater SNR must be demonstrated with that stimulus. In a trial of a novel stimulus form to be used for discriminating glaucoma from normality,

this can be done appropriately by comparing the quotient of the disease signal and response variability with that for the contemporary reference standard. In this experiment, a greater overall disease signal, lower dependence of response variability on depth of defect, and greater SNR, has been demonstrated with stimuli varying in area only, than for a standard Goldmann III stimulus, when compared on a common energy scale and with equivalent visibility. Mindful than an artefactual steepening of the psychometric function could be observed simply by using a psychometric function sampling protocol that enables a greater number of stimulus presentations, and larger range of energy levels, to be seen above threshold with one stimulus type in comparison to another (as discussed in chapter three), stimuli were matched for energy step size and spread of supra-threshold energy levels in the MOCS procedure. It was therefore ensured that any differences could reasonably be attributed to stimulus modulation, rather than an artefact of experimental design.

In *Figure 4.11*, a greater overall disease signal can be observed with all three test stimuli (A, AC and C_R) when compared with that for the standard GIII stimulus; the A stimulus showed a larger overall disease signal than the AC and C_R stimuli, as denoted by the greater overall TD (*Table 4.1*). As this was found with matched data, differing dynamic ranges between stimulus forms do not solely account for the differing disease signals observed here. The overall larger disease signal with the three test stimuli, compared with that for the GIII, is in keeping with the finding of a larger Ricco's area in glaucoma patients (displaced spatial summation curve along the area axis), relative to that for age-similar, healthy controls (Redmond et al. 2010a), as a difference in threshold to a GIII stimulus represents the distance, on the y-axis, between shallow regions of the spatial summation curves (*Figure 4.1*), for much of the central visual field. Threshold differences for the A, AC, and C_R stimuli, on the other hand, represent differences between glaucoma and normal curves in steeper regions of the curve.

Although the GIII stimulus used here has a slightly smaller area than that of a true Goldmann III, it is still beyond the area of complete spatial summation in the majority of healthy observers and those with early glaucomatous defects. From *Figure 4.4*, the GIII stimulus is expected to be beyond Ricco's area in 96.1 % of healthy, and 68.2 % of glaucoma participants at a visual field location of 10° eccentricity. In comparison, a

Goldmann III stimulus is expected to be beyond Ricco's area in 98.0 % of healthy, and 75.0 % of glaucoma participants at a visual field location of 10° eccentricity.

It could be assumed that the greater number of locations in which TD was not measurable with the C_R stimulus suggests that this stimulus is, in fact, superior at distinguishing between 'normal' and 'glaucoma' than the other three stimuli. However, the more likely explanation is that this reflects the smaller dynamic range for this stimulus with the hardware used in this experiment.

The greatest increase in response variability with depth of defect was found for the GIII stimulus (*Figure 4.13* and *Figure 4.14*). Response variability was found to be less dependent on disease severity for all three stimulus forms (A, AC, and C_R), with least dependence being observed with the A stimulus when all data were considered (complete data). Caution should be exercised at this point, however, as it does not necessarily follow that the stimulus with the lowest, or more uniform response variability has the greatest utility for disease detection. To answer this question, one must consider disease signal and response variability together.

Both the A and AC stimuli had a statistically significantly greater SNR than the GIII when all disease strata were considered, but this difference was greatest for the A stimulus. Although notable, the difference in SNR between the A and GIII stimuli is more modest in the lower stratum, likely due to these data representing not only locations with glaucomatous damage, but also those that are relatively healthy. By subdividing this stratum for further analysis, attempts were made to overcome this, however this action severely reduces the number of locations available for analysis, such that differences between stimulus forms become more difficult to ascertain with accuracy. Given that the SNR measure takes account of both disease signal and response variability, and that stimuli are compared on equivalent platforms, scales, and units, more substantial weight can be given to this metric in a comparison of their relative utility. In this experiment, tests were performed at locations [±7, ±7], i.e. 9.9° eccentricity from fixation. At this location, and at the background adaptation level employed, the area of the GIII stimulus is close to the normal Ricco's area. Further investigation in more central locations, using the methodology employed in this

experiment, may help to better understand the utility of area-modulated stimuli in the identification of earlier loss. It is noteworthy that the difference in SNR between the A and GIII stimuli is systematically enlarged across disease strata, while the difference in SNR between both the AC and C_R stimuli and the GIII stimulus remained modest (*Figure 4.18*). Although the utility of area-modulated stimuli for identifying change over time, and measuring remaining vision in advanced loss, was not formally investigated in this experiment, this finding raises the possibility that the A stimulus might also outperform the GIII in both regards.

Although some differences had been found in threshold and response variability in chapter three when the MOCS procedure was carried out on two different days, this did not appear to be the case in the small experiment carried out here (section *4.3.4*). No statistically significant difference was found between energy thresholds or response variabilities established on different days. Equally, no statistically significant difference was found between energy thresholds or response variabilities when the same tests were repeated twice in the same day, in the three healthy participants recruited. Firstly, this indicates that there is no significantly different fatigue effect with one stimulus form over another. As such, the results for those participants who elected to complete tests for two stimulus forms on one day are not expected to be substantially different from the results for those participants who elected to complete tests on four separate days. Secondly, although many external factors may influence test results that cannot be controlled for, the results do not appear unduly affected. This apparent improvement in repeatability in comparison with chapter three may be due to the additional staircase stage introduced at the beginning of the test, or may be due to the equidistant test locations used, aiding in the reduction of attentional bias.

The time interval between visits in this experiment is longer than is common in experiments of repeatability, which are typically conducted over eight weeks or less (Heijl et al. 1989a; Artes et al. 2002a; Wall et al. 2009). This is usually to ensure that no change in ocular status has occurred, and as such it could be argued that the time interval between visits in this examination of repeatability is too long. However, care was taken in this experiment to guard against such change, despite the time interval between visits, by confirming consistency of all ocular signs compared with those at

the initial visit. Indeed, it could be argued that this time interval is more meaningful in assessing repeatability, as it is more in keeping with a clinical setting, in which visual field tests are commonly conducted at intervals of at least six months. The lack of statistically significant differences in both energy threshold and response variability, confirmed that participants recruited to this experiment had not undergone any change in ocular status from the time of their initial visit.

Following reports of substantial retinal ganglion cell loss prior to clinical identification of glaucoma (Quigley et al. 1982; Kerrigan-Baumrind et al. 2000), possible vulnerability of ganglion cell subtypes to the condition (Quigley et al. 1987; Quigley et al. 1988; Dandona et al. 1991), and concerns about high variability in conventional clinical visual field testing (Henson et al. 2000; Artes et al. 2002a), the past few decades have observed a movement to establish test stimuli to identify, with high precision, the subtlest visual field damage. Many studies have previously attempted to compare the diagnostic capabilities of alternative tests with those of SAP. Such a comparison is not straightforward, however, and has been confounded by the use of differing apparatus, measurement scales, stimulus configurations, adaptation levels, and thresholding algorithms within and between studies. Firm conclusions about the superiority of one stimulus over another cannot be made without control over parameters outside those the manufacturer can provide on a clinical platform. A meaningful comparison of performance between different stimulus configurations can only be made if all other variables are accounted for or minimised. Therefore, caution should be exercised when making conclusions about the utility of one stimulus over another following a comparison on existing clinical platforms. Although it might initially be assumed that, when using clinical devices, one is comparing the effects of stimulus configuration alone, apparent differences in disease signal and response variability require more detailed consideration and explanation. For example, recent years have seen several investigations into the utility of the larger Goldmann V stimulus as a new reference standard for detecting glaucoma and identifying deterioration over time. If one wishes to compare the utility of the Goldmann III and V stimuli, it must be borne in mind that a 2 dB luminance increment for a Goldmann V stimulus is much larger in raw energy terms than a 2 dB luminance increment for a Goldmann III, e.g. $\log(\text{energy}_{GV,32dB} -$

energy$_{GV,30dB}$) > log(energy$_{GIII,32dB}$ – energy$_{GIII,30dB}$). When measuring the psychometric function, a larger spacing of stimulus levels in raw energy terms for the Goldmann V will result in a much more 'repeatable' threshold, in that the stimulus will be perceived to more definitively jump between 'seen' and 'not seen' around threshold; effectively an artefactual steepening of the psychometric function and thus an *apparently* lower response variability with a larger stimulus area. An apparent difference in response variability for stimuli of different area might therefore be conflated with the effects of unequal raw energy stimulus spacing. As such, it is important that a comparison of SNR with Goldmann III and V stimuli be made in which these confounders are accounted for, or at least minimised. By not limiting presented stimuli to integer dB steps, the experiment presented in chapter three (section *3.2*) minimised some of the confounding factors, but further investigation, in which equal raw energy steps are used between Goldmann stimuli, is recommended.

Redmond et al. (2010a) previously demonstrated that, in glaucoma, the difference in threshold from normal for a contrast-modulated stimulus close in area to a Goldmann III could be completely mapped to an enlarged Ricco's area (their Figure 5). It therefore follows that stimuli optimised to probe the change in spatial summation function in glaucoma may be more beneficial to identify subtle functional loss in early disease. The results of this experiment suggest that area-modulated stimuli may offer additional benefits for measuring glaucomatous changes in spatial summation in a clinical setting, in the form of greater disease signal, more uniform response variability with defect depth, and a greater SNR than the conventional fixed-area, contrast modulated stimuli (Goldmann III) currently employed in SAP.

The use of MOCS as a clinical test is not advocated; this design was chosen in order to ascertain the optimum stimulus modulation paradigm for probing changes in the visual field in glaucoma. Rather the utility of area-modulated stimuli should now be investigated further by comparison with conventional Goldmann III stimuli on an extended test grid, a common energy scale, common step sizes, and with a common thresholding algorithm, optimised for accuracy and test duration. In this way, the utility of these stimuli for the identification of visual field damage and its progression, can be confirmed in the clinical setting. The experiment presented in chapter five

takes this study one step further, examining a greater range of test locations with a slightly more 'clinic-friendly' protocol, and establishing test-retest variability characteristics with the same four stimulus forms.

Chapter 5 Test-retest variability of perimetric stimuli optimised to probe changing spatial summation in glaucoma

5.1   Introduction

As detailed in chapter four, the A stimulus was found to have a statistically significantly higher SNR compared with the conventionally used Goldmann III stimulus. Although MOCS is an essential 'first step' when investigating any new stimulus paradigm to permit investigation of response variability, it is only of use in a research setting as it affords more information than threshold alone. It is not an appropriate test to use in a clinical setting, due to the long test times and small numbers of test locations. Therefore, the logical 'second step' is to investigate whether a new stimulus paradigm translates to a clinical setting.

SAP, although employing various thresholding algorithms, e.g. SITA, to speed up testing time, is based on a standard staircase strategy ('method of limits'), permitting a greater number of visual field locations to be tested in a short time frame compared with MOCS. This method does not afford the necessary information to determine intra-test variability, as investigated in chapter four, but it is an appropriate method with which to assess test-retest variability of a stimulus paradigm. As previously discussed in section *1.3.3*, the test-retest variability of SAP is high in normal regions of the visual field, and in early disease, and increases with increasing disease severity (Wilensky and Joondeph 1984; Heijl et al. 1989a; Artes et al. 2002a). This results in two main disadvantages; poor discrimination of early disease from normal, and difficulties identifying disease progression from stable disease. The test-retest characteristics of the three novel stimulus forms (A, AC, and C_R) are, as yet, unknown.

Although intra-test variability with the alternative stimuli was noted to be more uniform than that for a Goldmann III equivalent stimulus (section *4.3.2*), it does not necessarily follow that a stimulus demonstrating consistent intra-test variability, will also benefit from consistent inter-test variability. In addition, Heijl et al. (1987; 1989a) noted that both intra- and inter-test variability were significantly greater in the mid-periphery than centrally. As the experiment presented in chapter four was conducted

within the central 10° only, it is necessary to investigate a wider range of eccentricities, to fully investigate the three novel stimulus forms (A, AC and C_R).

Two potentially confounding factors, in an investigation of inter-test variability, are learning effect and fatigue effect. Learning effect describes the phenomenon by which a greater perimetric sensitivity is observed on repetition of the test, i.e. as observers become more familiar with the test, their performance is seen to improve. The learning effect of SAP is well documented, and as a result any investigation into test-retest variability often ensures that participants have suitable perimetric experience (usually at least one perimetric result) prior to any experimental tests (Flammer et al. 1984; Wilensky and Joondeph 1984; Heijl et al. 1989a; Artes et al. 2002a). However, studies that have specifically investigated the learning effect note that there is considerable variation between observers (Heijl et al. 1989b), apparent differences in learning effect between healthy and glaucomatous observers (Heijl et al. 1989b; Wild et al. 1989; Heijl and Bengtsson 1996), and differences across the visual field (Heijl et al. 1989b; Heijl and Bengtsson 1996). Fatigue effect operates in the opposite manner to learning effect, whereby perimetric sensitivity is observed to decline, and has been reported both intra- and inter- test (Wild et al. 1989). The length of the test, influenced by the number of visual field locations being tested, and the method by which sensitivity is established, may influence the fatigue effect. Design of a staircase testing strategy often involves a balance between the number of test locations, and the number of reversals at each location, to achieve a sensitivity at an acceptable number of visual field locations within an acceptable time frame. A greater number of reversals will generally increase the accuracy of the determined threshold, but will also result in a longer test time, and therefore a greater fatigue effect, which may decrease the accuracy of thresholds determined with time. A test with a greater intra-test fatigue effect will also influence test-retest variability.

Test-retest variability has also been found to differ between healthy and glaucomatous observers (Heijl et al. 1989a; Artes et al. 2002a). Several studies note both an 'improvement' and a 'deterioration' which could feasibly be interpreted as a 'learning' or 'fatigue' effect respectively, but could equally be interpreted as random variation (Katz and Sommer 1987; Wild et al. 1989; Wild et al. 1991). Indeed, Heijl et al.

(1989b), in their investigation of learning effect, noted that in the majority of healthy, perimetrically naïve observers, results improved little or not at all over multiple tests, with an average sensitivity increase of 1.3 dB between the first few tests. In their study investigating the learning effect in those with suspect glaucoma and early glaucoma, Gardiner et al. (2008) noted an improved mean sensitivity of 0.5 dB over the first year, then no further improvement. Other studies inform us that test-retest variability is often higher than this (Wilensky and Joondeph 1984; Artes et al. 2002a), although these values were determined point-wise, rather than as an average sensitivity. As such, and as noted by Werner et al. (1988), it can be difficult to untangle the learning and fatigue effects from inherent random variation.

Other factors which may influence test-retest variability include the position of visual field locations, and the order in which these locations are tested. Although there is some randomisation of stimulus presentation to prevent observers predicting the location of the next stimulus, this is not a complete randomisation of all test locations. In many thresholding algorithms, e.g. SITA, the perimetric sensitivity of four 'seed' locations within the central 13° are initially determined. From these sensitivity values, those of neighbouring locations may be more accurately estimated, and the intensity of subsequently presented stimuli are tailored to the observer's expected sensitivity, permitting a quicker termination of the strategy at each location (Bengtsson et al. 1997a; Bengtsson and Heijl 1998a; Bengtsson and Heijl 1998b; Bengtsson et al. 1998). Presented locations therefore become generally more peripheral as the test progresses, which can result in the so-called 'cloverleaf' field plot in those observers whose attention wanes with test duration (Heijl et al. 2012). Test-retest variability has been found to vary with eccentricity (Katz and Sommer 1986; Heijl et al. 1987; Heijl et al. 1989a; Heijl et al. 1989b; Chauhan and House 1991), which could be due to a fatigue effect induced by this test pattern.

The purpose of this experiment was to investigate the test-retest variability of the same four stimulus paradigms investigated in chapter four, with the hypothesis that stimuli optimised to probe the differing spatial summation between healthy and glaucomatous test locations would demonstrate a lower test-retest variability. Steps were taken to control for potential learning or fatigue effect that may mask this

variability. All participants had experience with SAP, although it should be noted that studies have indicated that experience with one form of perimetry does not necessarily translate to other forms of perimetry (Wild and Moss 1996; Gardiner et al. 2008). Some participants had previously taken part in the experiment presented in chapter four, and therefore had experience with each of the three novel stimulus forms (A, AC and C_R), while other participants were naïve to these stimuli. A comparison between these groups determined whether a learning effect could be observed. Fatigue effect was controlled for by randomisation of stimulus order for each participant on each day.

Commercially available perimeters utilise a thresholding algorithm, often based on maximum-likelihood principles, such that perimetric sensitivity may be determined at a greater number of test locations than would otherwise be possible, whilst maintaining an acceptable test time. In examination of the test-retest variability, as proposed here, it is desirable to mimic a visual field test that might be used in a clinical setting. As such, a greater number of test locations are required than used previously in chapter four. The potential advantage in using a thresholding algorithm is that a greater number of test locations could be utilised, permitting investigation of the widest range of eccentricities and perimetric sensitivity values for the recruited cohort. However, the use of a thresholding algorithm does influence the sensitivity outcome, as evidenced by Artes et al. (2002a), in which test-retest variability was investigated between Full Threshold, SITA Fast and SITA Standard strategies. Although sensitivity values and test-retest variability were similar between these strategies, they were not identical. For a perimetric sensitivity greater than 20 dB, the 90% retest intervals of SITA Standard and SITA Fast were smaller than that of the Full Threshold strategy, suggesting that sensitivity values were more repeatable with these algorithms. However, for a perimetric sensitivity less than 20 dB, the 90% retest intervals were larger with the SITA Fast strategy, and similar between the SITA Standard and Full Threshold strategies. The Full Threshold strategy uses a fixed 4-2-2 dB staircase procedure, whereas both SITA strategies, while based on a staircase procedure continually estimate sensitivity and measurement errors, based on observer responses to the presented stimuli (Bengtsson et al. 1997a; Bengtsson and Heijl 1998a; Bengtsson

and Heijl 1998b; Bengtsson et al. 1998), therefore step size is likely not fixed. As a dB increment is larger in raw energy terms at the lower end of the dB scale, this difference in 90% retest intervals could be explained by a higher effective step size for SITA strategies at threshold sensitivities greater than 20 dB. In addition, given that all perimetric thresholding algorithms have been designed for use with a Goldmann III stimulus, this may create a possible bias when comparing with other stimulus paradigms. Therefore, although the use of a thresholding algorithm would be essential in clinical practice, the decision was made to use a basic, adaptive staircase strategy for the purposes of this experiment. Thus, any differences found in test-retest variability may be reliably attributed to the stimulus itself, and are not confounded by effects of the thresholding algorithm.

Given that the number, and position of test locations may influence test-retest variability, these were carefully considered in the study design, as detailed in section *5.2*. As these factors, in addition to the use of a thresholding algorithm, may influence sensitivity, the test-retest variability findings reported in the studies cited above may not be directly applicable. Test-retest variability of any of the four stimulus forms, including that of the Goldmann III equivalent (GIII), may be found to differ substantially from that previously reported.

5.2 Methods

In this cross-sectional study, thresholds were established with each of the four stimulus forms (A, AC, C_R, and GIII) at 18 test locations, in participants with glaucoma and age-similar controls, using an adaptive, 1:1 staircase strategy. This was repeated five times, over five visits, for each stimulus form. The 5th and 95th retest percentiles, were compared between stimulus forms.

5.2.1 Participants

Fifteen participants with glaucoma (median [IQR] age: 69.5 years [67.5, 72.1] and five healthy, age-similar participants (median age: 71.9 years [67.6, 72.5]) were recruited to this experiment. All of the participants with glaucoma had received a diagnosis of POAG, six with high tension, and nine with normal tension glaucoma, by the hospital eye service. As the aim of this experiment was to determine test-retest characteristics

of the four stimulus forms for a wide range of sensitivities, all glaucoma participants had a repeatable defect in their visual field, ranging in severity from near normal ('borderline' on the Glaucoma Hemifield Test) to 'advanced' field loss (categorised as per the Hodapp-Parrish-Anderson grading scale; Hodapp et al. 1993), with the SITA Standard 24-2 program on the Humphrey Field Analyzer (HFA II, Carl Zeiss Meditec Inc., Dublin, CA). Median [IQR] MD was -4.88 dB [-6.78, -2.62]. All healthy participants had a normal visual field ('within normal limits') with the SITA Standard 24-2 program on the HFA II; median [IQR] MD was +0.46 [-1.26, +0.92].

Only one eye of each participant was tested; this eye was selected as the eye that best met the inclusion/exclusion criteria, or was selected at random if both eyes were equally suitable. Of the participants with glaucoma, five right eyes and ten left eyes were tested, and of the healthy participants, three right eyes and two left eyes were tested. SAP (HFA II, SITA Standard 24-2 program) was performed twice in the test eye prior to any experimental tests, or once if participants had undertaken one of these tests within the past six months as part of their routine clinical care. False positive rates were < 15% for all participants. In addition to perimetric experience with the HFA II, nine of the fifteen participants with glaucoma, and three of the five healthy participants, had previously taken part in the experiment presented in chapter four; as such, these participants had prior experience with the three novel stimulus forms. Each participant attended on five separate visits. Tests were completed for each of the four stimulus forms at each visit, conducted in a random order at each visit to counteract any fatigue effect. Participants completed all experimental tests within an eleven-week period.

The same inclusion/exclusion criteria as those detailed in chapter four applied in this experiment. Participants did not have any other ocular/systemic disease and/or medication known to affect visual performance (e.g. diabetes, thyroid disease, age-related macular degeneration, hydroxychloroquine medication); ocular health was confirmed by slit lamp biomicroscopy at each visit. One participant with glaucoma had previously undergone trabeculectomy surgery in the test eye seven years prior to the study; this eye had been considered stable since the surgery. Otherwise, no participants had undergone any ocular surgery, with the exception of uncomplicated

cataract surgery. All participants had an IOP < 21 mmHg at each visit, measured with Goldmann Applanation Tonometry. Healthy participants did not have any first-degree relatives with glaucoma, and did not have a history of elevated IOP.

All participants had a best-corrected visual acuity of ≥ 6/9 (confirmed at all visits), in the absence of significant corneal or media opacities (≤ NO3, NC3, C3, and/or P3, Lens Opacities Classification System III; Chylack et al. 1993), with a spherical refractive error between +6.00 DS and -6.50 DS, and astigmatism < 3.50 DC in the test eye, as determined by a full refraction conducted before the commencement of any experimental tests. In those participants who had previously undergone cataract surgery, pre-surgical refractive errors that did not meet these criteria were also excluded (if known). A half-eye trial frame was worn, and full aperture trial lenses were used to correct refractive error for a viewing distance of 30 cm. All experimental tests were conducted with natural pupils, and the non-test eye was occluded with a patch.

Ethical approval for the study was given by the East of Scotland Research Ethics Committee (NHS Scotland). The research adhered to the tenets of the Declaration of Helsinki. Written, informed consent was obtained from all participants prior to inclusion.

5.2.2   Apparatus and set-up

The same apparatus was used as that described in chapter two. Stimuli were displayed on a gamma-corrected, 25" OLED display (Sony PVM-A250 Trimaster El, resolution 1920 x 1080 pixels, frame rate 60 Hz, refresh rate 120 Hz), driven by a ViSaGe MKII Stimulus Generator (Cambridge Research Systems, Rochester, UK). Experiments were programmed in MATLAB (version 2014b; The MathWorks, Inc., Natick, MA) using the CRS toolbox (version 1.27, Cambridge Research Systems, Rochester, UK), and adapted from the staircase procedure used in chapter four (section *4.2.4.1*). A nominally uniform background luminance of 10 cd/m² was used. During all tests, participants were instructed to fixate a central cross on the screen, and respond to any stimulus they detected in their visual field by pressing a button on a response pad (Cedrus RB-530; Cedrus, USA).

Parameters for the AC, C_R, and GIII stimulus forms were the same as described in chapter four (section *4.2.3*); the A stimulus differed slightly, with the maximum area capped at 25.48 deg^2 (5.68° diameter), thus preventing overlap with any adjacent test locations in a standard 24-2 grid pattern, in which test locations are at 6° intervals. Stimulus duration was fixed at 0.2 seconds in all experiments.

5.2.2.1 Test locations

The number of test locations used was informed by the staircase strategy undertaken by participants in chapter four. The average time taken for this staircase task was determined for healthy observers and those with glaucoma. From these test times, it was established that a test consisting of 18 test locations, with four reversals at each location, plus the inclusion of eight false positive and eight false negative catch trials, should not exceed the typical test duration of a SITA Standard strategy.

Four of the locations used here were the four test locations used in chapter four; these locations are not from a standard 24-2 test pattern. Fourteen additional test locations were established, guided by the work of Wang and Henson (2013), who investigated sub-sets of the standard 24-2 test grid of differing numbers of test locations, and identified those test patterns which identified the greatest number of visual field defects. Ten locations of the locations selected here were those of the optimal test pattern identified by Wang and Henson (2013) for a subset of ten test locations. Four additional locations were selected from the optimal test pattern identified by Wang and Henson (2013) for a subset of 20 test locations, such that there were equal numbers of test points in the superior and inferior hemifields. These 18 test locations are shown in *Figure 5.1* for a right eye; the mirror image of these locations was used when testing a left eye.
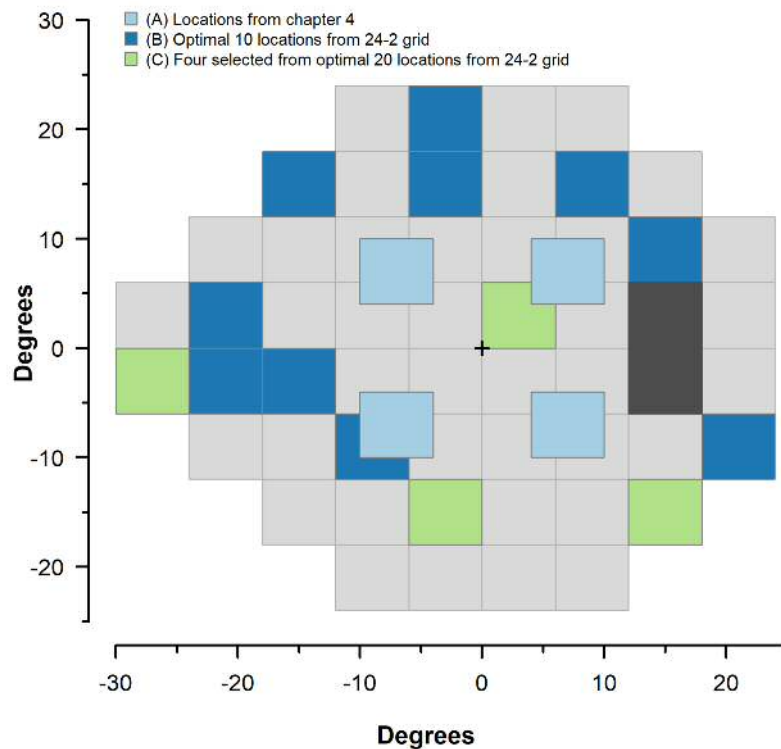
Figure 5.1 – Test locations used in this experiment (shown for a right eye). Locations informed by the work of Wang and Henson (2013).

5.2.2.2 Determining the initial stimulus energy at each test location

An appropriate 'initial stimulus', i.e. the first stimulus to be presented at each of the 18 test locations, was established from the findings of chapter four. The initial stimulus at the four locations tested in chapter four, (A) in *Figure 5.1*, was established first for each of the four stimulus forms, as described below; this will be referred to as IS_O. The initial stimulus for the 14 supplementary locations, (B) and (C) in *Figure 5.1*, were then established by applying an energy adjustment per eccentricity, informed by the findings of Khuu and Kalloniatis (2015b); this will be referred to as IS_S.

Initial stimulus for the original four locations tested in chapter four (IS_O)

Figure 5.2 shows the mixed model linear regression from section *4.3*, which modelled the relationship between threshold energy, pooled for the original four locations tested, and age in healthy participants. Threshold energy was found to increase by ~0.05 log energy per decade, the minimum possible threshold difference set by the parameters of the study, with slight differences noted between stimuli. IS_O was

determined as 0.5 log energy above the expected threshold per decade of age, established for each stimulus form from the linear regression model, as is commonly used in supra-threshold perimetry (Spry et al. 2000; Artes et al. 2002b; Artes et al. 2003a).
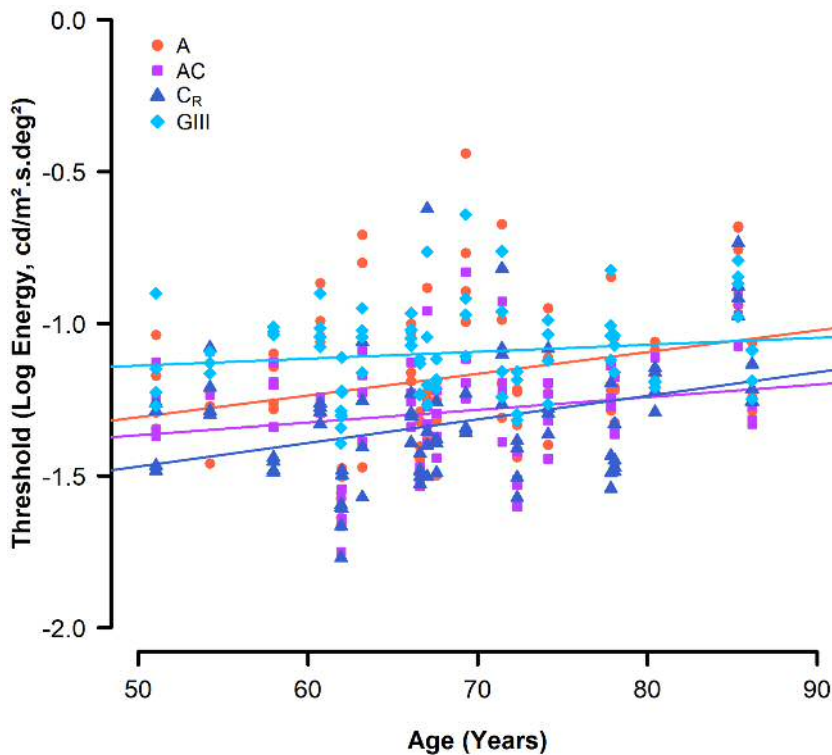
Figure 5.2 – Threshold energy, pooled for the four locations tested in chapter four, plotted against age and fitted with a mixed model linear regression.

Initial stimulus for the 14 supplementary locations (IS_S)

IS_S was determined by applying an energy adjustment per eccentricity to IS_O. This was informed by the findings of Khuu and Kalloniatis (2015b), who investigated the difference in perimetric sensitivity with every 10° from fixation, to a maximum of 30°, for each of the five Goldmann stimuli in healthy observers.

To determine the energy adjustment for the GIII stimulus, the perimetric sensitivity values established by Khuu and Kalloniatis (2015b) for a Goldmann III stimulus were determined along the 45-225° and 135-315° meridians for the eccentricities of each test location used in this experiment (between 4° and 27° from fixation, *Figure 5.1*).

Each perimetric sensitivity value was converted from dB to stimulus energy, in accordance with *Equation 3.3*, and averaged per eccentricity. The differences between the energy value for a Goldmann III stimulus at 10° from fixation, (A) in *Figure 5.1*, and a Goldmann III stimulus at each eccentricity for the other 14 test locations, (B) and (C) in *Figure 5.1*, were determined. The appropriate energy adjustment for the 14 supplementary test locations, according to eccentricity, were then applied to IS_O for the GIII stimulus, to determine IS_S; the initial stimulus for both IS_O and IS_S was based on eccentricity only, and did not differ according to quadrant.

To determine IS_S for the C_R stimulus (which has an area greater than Goldmann I, but less than Goldmann II), the same process was repeated for the Goldmann I and Goldmann II stimuli in Khuu and Kalloniatis (2015b). Findings were averaged for these two stimuli before applying the energy adjustment per eccentricity to IS_O, to give IS_S for the 14 supplementary test locations. As the energy steps were equivalent between the A, AC, and C_R stimuli, as discussed in *4.2.3*, the same energy adjustment as C_R was applied to IS_O for A and AC, to give IS_S for the 14 supplementary test locations.

5.2.2.3 Test procedure

An adaptive, 1:1 staircase procedure of four reversals was conducted at each of the 18 test locations (randomly interleaved). The stimulus increased/decreased in energy by 0.5 log units until the first reversal, 0.25 log units until the second reversal, 0.1 log units until the third reversal, and 0.05 log units until the fourth, and final, reversal. Threshold at each location was determined as the mean of the final two reversals (one ascending, and one descending). For those test locations at which the true threshold lay beyond the maximum log energy available, the threshold was recorded as the maximum possible log energy value for that stimulus form (similar to a SAP sensitivity recorded as '0 dB' when perimetric sensitivity lies beyond the maximum luminance range of the perimeter). Eight false positive and eight false negative catch trials were also included. At eight random occurrences throughout the test no stimulus was presented, and a positive response from the observer was recorded as a false positive. At eight random occurrences throughout the test, a stimulus was presented at 1.5 log energy higher than the first reversal at that test location (the stimulus energy used to determine p(seen, 1.0) in chapter four); if the observer failed to respond, this was

recorded as a false negative. Tests in which the false positive rate was higher than 25% (more than two false positive responses) were discarded, and the test was repeated. In addition to monitoring observer inattention, the use of false negative catch trials can promote observer attention, as the presented stimulus should be substantially supra-threshold, and therefore more easily detected than those near threshold, such that observers respond with a higher level of certainty. Tests in which the false negative rate was high were not repeated, as high false negative rates in a glaucomatous visual field test are often not an indication of inattention, but due to the disease itself (Bengtsson and Heijl 2000).

Participants completed one test for each of the four stimulus forms at each visit, with the order of tests randomised at each visit, and a rest break taken between tests. Participants could also request additional rest breaks as required.

5.2.3   Statistical analysis

Statistical analyses were performed with the freely available, open source statistical environment R (R Development Core Team, 2017), and SPSS (IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0, Armonk, NY: IBM Corp).

5.2.3.1 Learning/fatigue effect

To investigate the learning/fatigue effect of the four stimulus forms, participants were divided into two categories – those who had taken part in the experiment presented in chapter four (twelve participants in total, nine with glaucoma and three healthy), and those who had not (eight participants, six with glaucoma and two healthy). Participants who had taken part in the experiment presented in chapter four had prior experience of the A, AC, and C_R stimuli ('experienced participants'), while other participants had experience of a Goldmann III stimulus, but were inexperienced with the other three stimulus forms ('novice participants'). Mean thresholds for the 18 test locations were calculated for each test undertaken by each participant. A one-way ANOVA with repeated measures was conducted to compare the effect of test number on mean threshold. This was conducted individually for each stimulus form, and separately for experienced and novice participants. Mauchly's Test of Sphericity was

used to determine whether assumptions of sphericity were violated; where this assumption was violated, a Greenhouse-Geisser correction was applied.

Statistical comparisons were not made between experienced and novice groups, as disease severity differed between these groups. As such, any statistical difference between mean thresholds would likely reflect this, rather than differences in learning/fatigue effect.

5.2.3.2 Test-retest variability

<u>Full dataset</u>

As the five tests were conducted over a short period of time, there is no one test that represents a true 'baseline', as no change is expected to have occurred during this time-period. As such, all five possible permutations of test order were considered for each stimulus form, with the first test of each permutation established as the 'test' threshold, and the remaining four tests established as the 'retest' thresholds; these five combinations were then collated and displayed as a box-and-whisker plot with the 'test' values shown on the x-axis, and 'retest' values shown on the y-axis, in a similar manner to that of Artes et al. (2002a). Thus, each plot represented all five permutations of test order. This approach helped to control for any fatigue effect that may occur from completing tests for each of the four stimulus forms in one session. Test thresholds along the x-axis were grouped in 0.1 log energy intervals, and 90% retest intervals (between the 5th and 95th percentiles) were determined. Threshold values for each of the 18 test locations were included. Plots were constructed incorporating all glaucoma and healthy participants together, in addition to considering healthy participants separately.

To determine which stimulus form displayed the lowest test-retest variability, the area between the 5th and 95th retest percentiles was determined for each stimulus. To ensure an appropriate comparison between all four stimulus forms, maximum thresholds for the A, AC and GIII stimuli were truncated to match that of the C_R stimulus, which had the lowest dynamic range in this experiment, enabling a true comparison between all stimuli. To compare test-retest variability beyond the

maximum threshold for the C_R stimulus, comparisons were made between the A, AC and GIII stimuli only, as described in section *5.3.2*.

<u>Subdivision of data into complete and incomplete spatial summation</u>

In addition to examining all test locations together as described above, test locations which underwent complete spatial summation with a Goldmann III stimulus were considered separately from those which underwent incomplete spatial summation with a Goldmann III stimulus, according to the two-phase hockey-stick model devised by Swanson et al. (2004); this model describes the relationship between log retinal ganglion cell count and perimetric sensitivity. Swanson et al. (2004) noted that, for a Goldmann III stimulus, perimetric sensitivity was determined by the area of this stimulus in relation to its position on the spatial summation curve, i.e. whether this stimulus underwent complete or incomplete spatial summation. By plotting perimetric sensitivity (dB) by retinal ganglion cell number, they noted a 'break point' where a linear regression fit of the data changed gradient, resulting in their two-phase model, a schematic of which is illustrated in *Figure 5.3*. The break point represented Ricco's area, the limit of complete spatial summation, which equated to a log ganglion cell count of 1.49, and a perimetric sensitivity of 31 dB. This two-phase model accounted for 82% of the data in Swanson et al. (2004).

The perimetric sensitivity data used to construct this model were based on data collected by Heijl et al. (1987) from healthy observers, using a Humphrey Field Analyzer and a Full Threshold strategy, with a 30-2 test grid. As such, it cannot be directly applied in this experiment, given that participants were tested using the SITA Standard strategy with a 24-2 test grid; as noted in section *5.1*, the thresholding algorithm used may influence the resulting perimetric sensitivity value. Bengtsson et al. (1998) and Bengtsson and Heijl (1998a) noted that, due to the methods by which differing strategies determine perimetric sensitivity, the expected difference between a Full Threshold and a SITA Standard strategy was 1 dB, with the SITA algorithm expected to give the higher estimate of perimetric sensitivity. While this expected difference between strategies may not always be exact, it has been found to be a close estimate (Bengtsson and Heijl 1998a; Bengtsson and Heijl 1998b; Artes et al. 2002a). A modified perimetric sensitivity was therefore determined by subtracting 1 dB from the

perimetric sensitivity value for each test location, established from the SITA Standard 24-2 program (HFA II, conducted prior to the experimental tests). For the four locations tested in chapter four, (A) in *Figure 5.1*, perimetric sensitivity values were predicted using bi-linear interpolation with MATLAB (version R2015b; The MathWorks Inc., Natick, MA, USA). These modified sensitivity values were used to establish the position of the test location on the spatial summation curve, in accordance with the two-phase model of Swanson et al. (2004). Although differing test grids were also used, which may impact estimates of perimetric sensitivity, this was deemed an acceptable method.
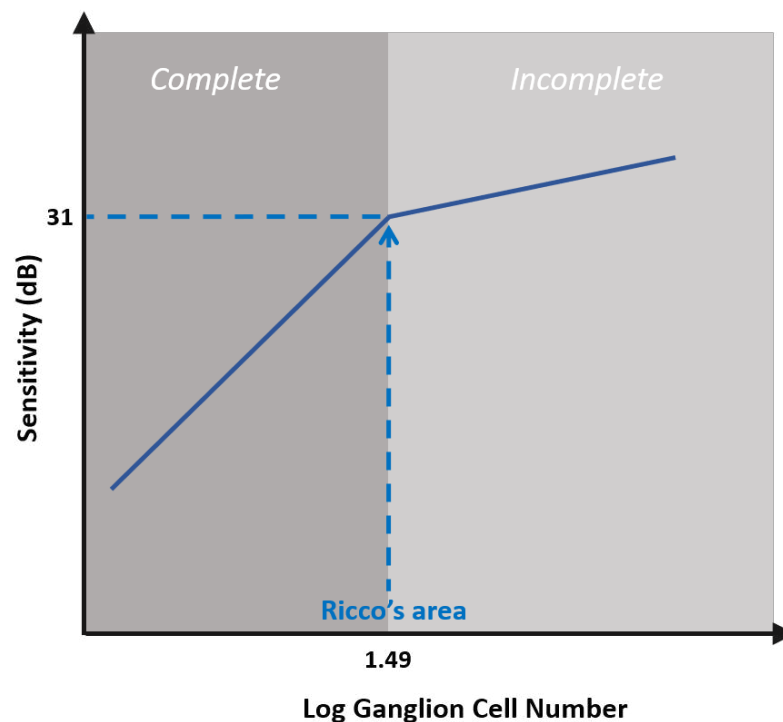
Figure 5.3 – Schematic illustrating the two-phase hockey-stick model of Swanson et al. (2004), describing the relationship between log retinal ganglion cell number and perimetric sensitivity (dB). Sensitivities ≤ 31 dB are within the area of complete spatial summation for a Goldmann III stimulus. Those > 31 dB are beyond Ricco's area, and therefore undergo incomplete spatial summation.

Test locations were then subdivided into two categories; those locations with a modified perimetric sensitivity ≤ 31 dB, at which a Goldmann III stimulus would undergo complete spatial summation (CSS), and those locations with a modified

perimetric sensitivity > 31 dB, at which a Goldmann III stimulus would undergo incomplete spatial summation (ISS). As for the full dataset, all five permutations of test order were determined, and the 5th and 95th retest percentiles for each stimulus form were established. The area between these percentiles was then determined.

5.3   Results

As a clinical indicator of the range of defect depth at the tested locations in this experiment, histograms indicating the raw dB and TD values from the final preliminary SAP test (HFA II SITA Standard 24-2) for the 18 test locations are shown in *Figure 5.4*. Values for the four test locations used in chapter four were predicted using bi-linear interpolation with MATLAB, as described in section *4.3*.
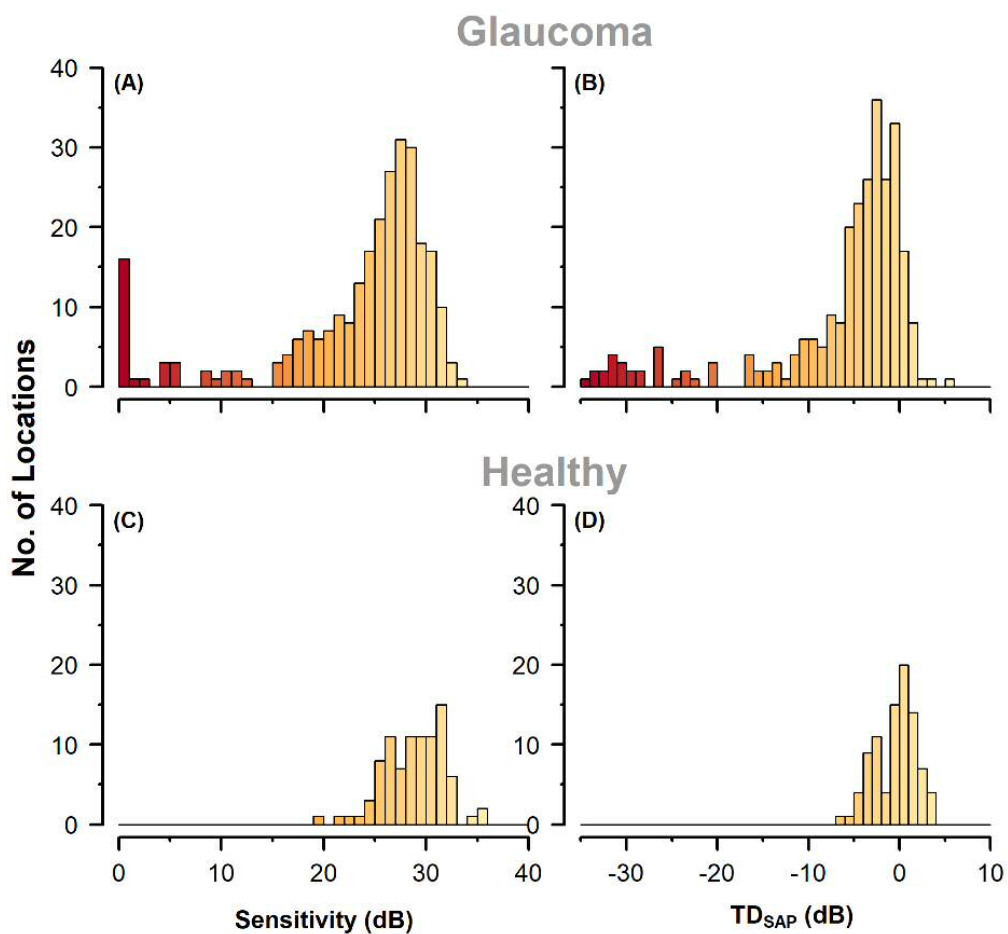
Figure 5.4 – Perimetric sensitivity (HFA II, SITA Standard 24-2), for the 18 test locations used in this experiment. (A) Raw dB values for participants with glaucoma. (B) TD values for participants with glaucoma. (C) Raw dB values for healthy participants. (D) TD values for healthy participants.

5.3.1   Learning/fatigue effect

Box-and-whisker plots of mean threshold of the 18 test locations, from each of the five tests, with each of the four stimulus forms, are shown in *Figure 5.5* (experienced participants), and *Figure 5.6* (novice participants). Healthy participants, and those with glaucoma, are included in these plots.
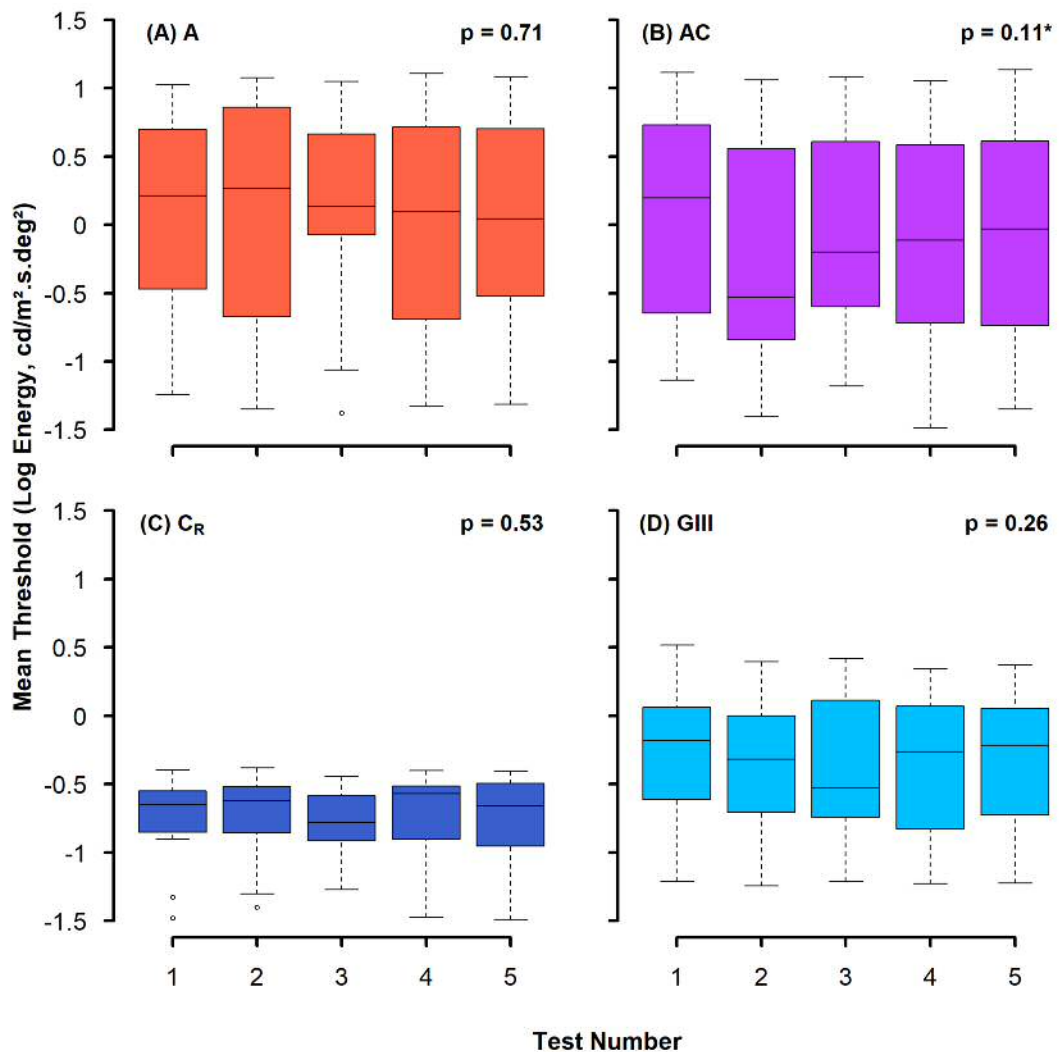
Figure 5.5 – Box-and-whisker plots for each of the five tests, for participants who had prior experience of the four stimulus forms (A-D). P-values are displayed from the one-way ANOVA. *Greenhouse-Geisser correction applied.
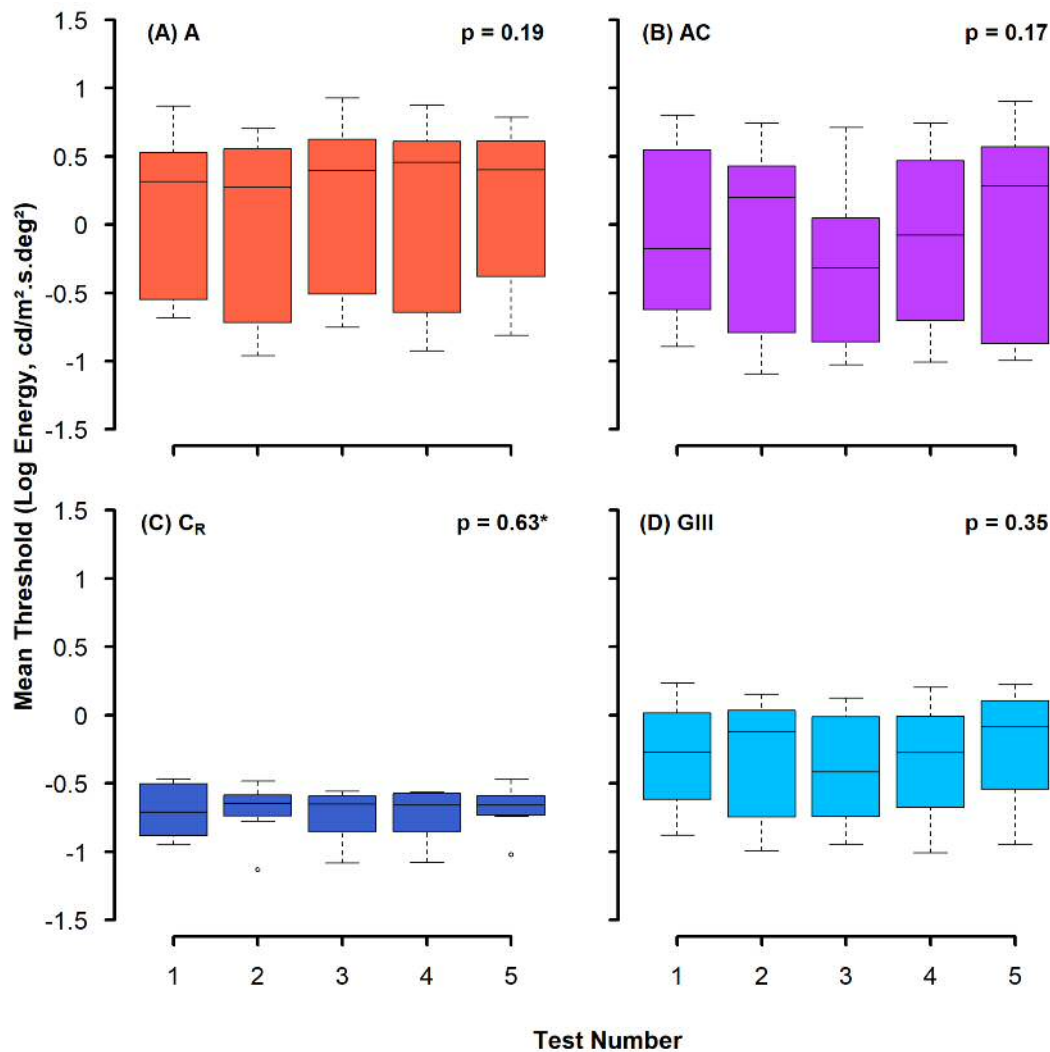
Figure 5.6 – Box-and-whisker plots for each of the five tests, for participants who had prior experience of a Goldmann III stimulus (D) but were inexperienced with the other three stimulus forms (A-C). P-values are indicated from the one-way ANOVA. *Greenhouse-Geisser correction applied.

A substantial learning effect would be indicated by a notable reduction in mean threshold with increasing test number, while a substantial fatigue effect would be indicated by a notable increase in mean threshold with increasing test number. For both experienced and novice participants, median values for mean threshold of A and C_R stimuli appeared uniform across the five tests (*Figure 5.5* and *Figure 5.6, A & C*). The range of mean threshold values was much narrower for the C_R stimulus than for the other three stimulus forms, which may be due to the shorter dynamic range of this stimulus with the apparatus used here (maximum available energy: -0.33 cd/m^2.s.deg^2). For the GIII, and particularly the AC stimuli, median values were more

variable across the five tests (*Figure 5.5* and *Figure 5.6*, *B & D*), but did not appear to indicate a particular learning or fatigue effect.

The range of mean threshold values was perhaps narrower in novice participants than experienced participants across all four stimulus forms. It is likely that this represents the differing threshold ranges between the two groups.

A one-way ANOVA with repeated measures found no statistically significant difference between tests for any of the four stimulus forms, for both experienced and novice participants; p-values from these analyses are displayed in *Figure 5.5* and *Figure 5.6*. Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated when testing the AC stimulus in experienced participants (p = 0.008) and when testing the C_R stimulus in novice participants (p = 0.039); the p-value shown for these stimuli has a Greenhouse-Geisser correction applied, as indicated by * in *Figure 5.5* and *Figure 5.6*.

As no discernible learning/fatigue effect was noted with any of the four stimulus forms across the five tests, and as no discernible difference was noted between experienced and novice participants, all five tests were included in further analysis, and data from all participants were grouped together from this point onwards.

5.3.2   Test-retest variability

 5.3.2.1 Full dataset

Box-and-whisker plots

As described in section *5.2.3.2*, test and retest data for all five permutations of test order were collated for each stimulus form, and plotted as box-and-whisker plots of retest against test. *Figure 5.7* shows these bow-and-whisker plots for all 18 test locations, incorporating both glaucoma and healthy participants. Test thresholds have been grouped in 0.1 log energy intervals, and the 5th and 95th retest percentiles are indicated by the solid lines.

In observing *Figure 5.7*, all four stimulus forms appear to show a greater test-retest variability with increasing threshold (reduced sensitivity), in keeping with Artes et al. (2002a). The flattened shape at the top of each plot indicates a ceiling effect, whereby

the maximum log energy for that stimulus form was been reached. With the apparatus used in this experiment, the C_R stimulus had the shortest dynamic range, reaching its maximum threshold first (*Figure 5.7.C*), followed by the GIII stimulus (*Figure 5.7.D*). The A (*Figure 5.7.A*) and AC (*Figure 5.7.B*) stimulus forms had very similar dynamic ranges in this experiment, although the AC stimulus had a slightly larger range.
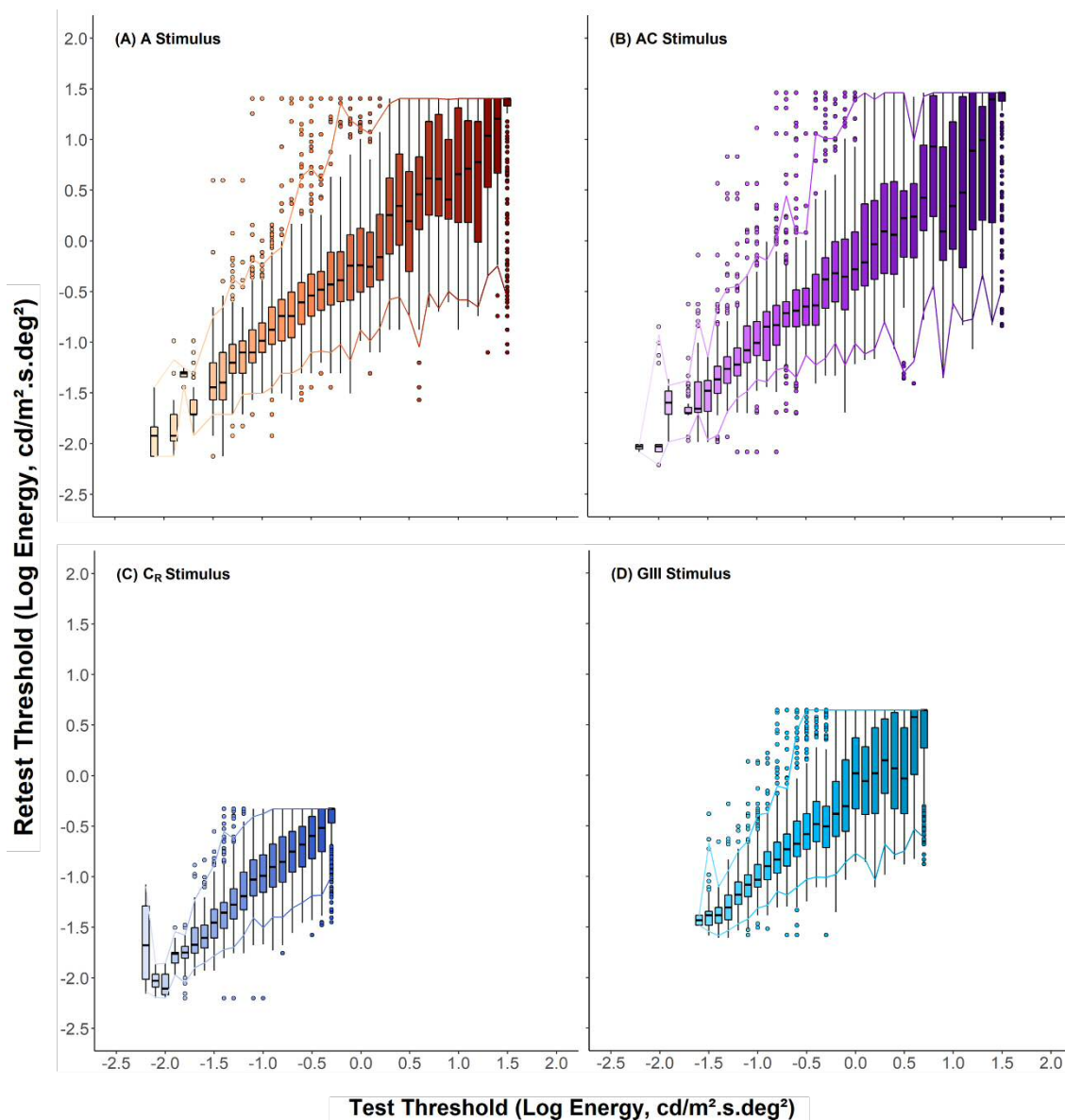
Figure 5.7 – Retest plotted against test thresholds (grouped in 0.1 log energy intervals) for the 18 test locations, for all four stimulus forms, incorporating thresholds from both glaucoma and healthy participants. Solid lines indicate the 5th and 95th retest percentiles.

Figure 5.8 shows box-and-whisker plots for all 18 test locations for healthy participants only. Observation of *Figure 5.8* suggests that those stimuli modulating in contrast only (C$_R$ and GIII, *Figure 5.8.C & D*) appear to show a greater variability with increased threshold, while those stimuli modulating in area (A and AC, *Figure 5.8.A & B*) appear to show a greater uniformity with increased threshold. However, as there are more outliers with the area modulating stimuli than the contrast-only modulating stimuli, accurate analysis of the 5th and 95th retest percentiles is difficult.
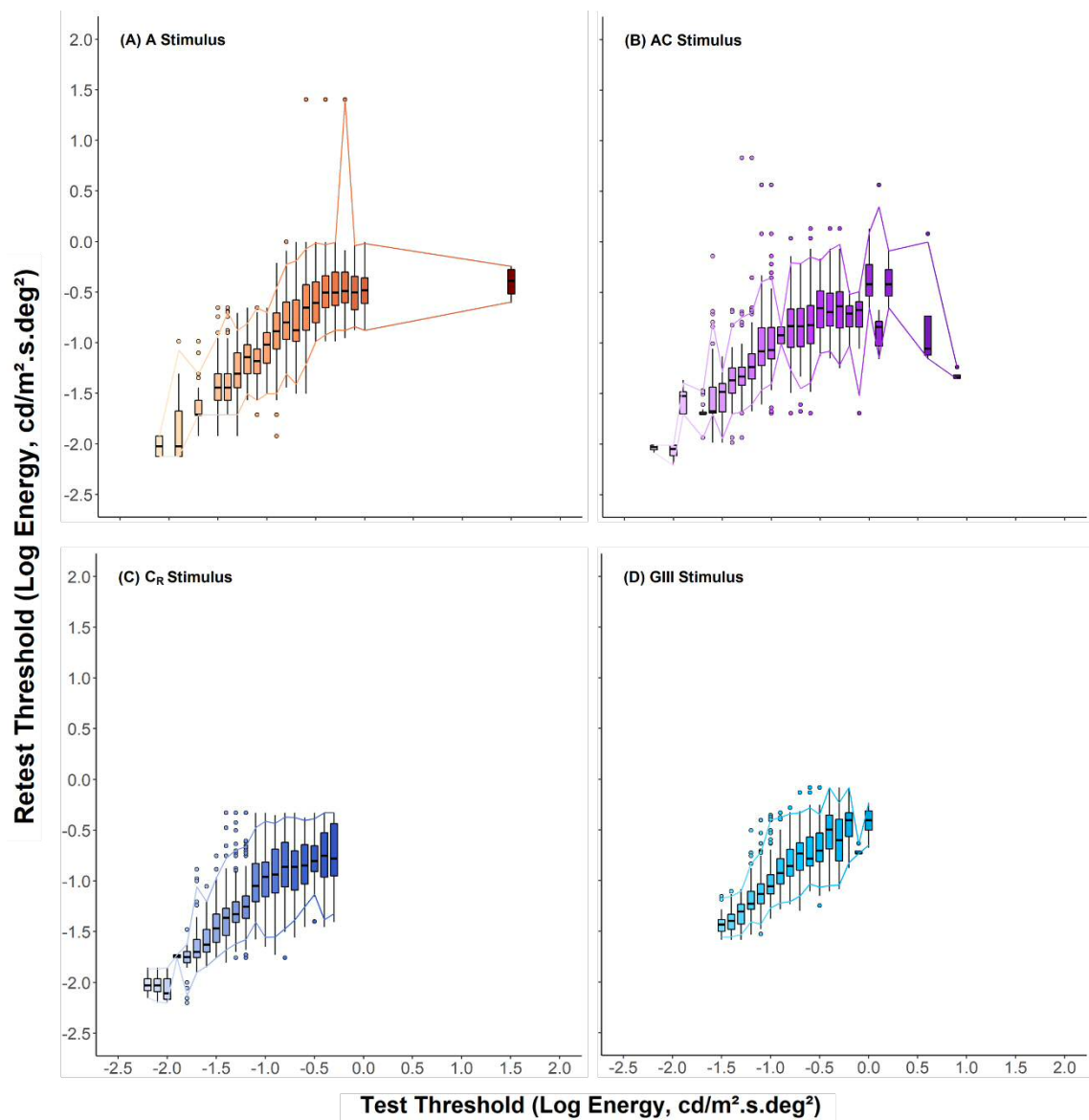
Figure 5.8 – Retest plotted against test thresholds (grouped in 0.1 log energy intervals) for the 18 test locations for the four stimulus forms, incorporating thresholds from healthy participants only. Solid lines indicate the 5th and 95th retest percentiles.

Figure 5.9 shows box-and-whisker plots for the four locations tested in chapter four, (A) in *Figure 5.1*, incorporating both glaucoma and healthy participants. As *Figure 5.8* incorporates healthy participants only, lower thresholds generally denote those test locations closer to fixation, while higher thresholds denote those test locations further from fixation. As the locations in *Figure 5.9* were equidistant from fixation, lower thresholds generally denote near-normal test locations, while higher thresholds denote more damaged test locations. As in *Figure 5.7*, a greater test-retest variability is noted with an increase in threshold for all four stimulus forms. This may indicate that test-retest variability for area-modulating stimuli (A and AC) is greater with increasing threshold due to glaucomatous damage, but remains more uniform with increasing threshold due to increasing eccentricity from fixation, although as previously noted it is difficult to analyse this accurately.
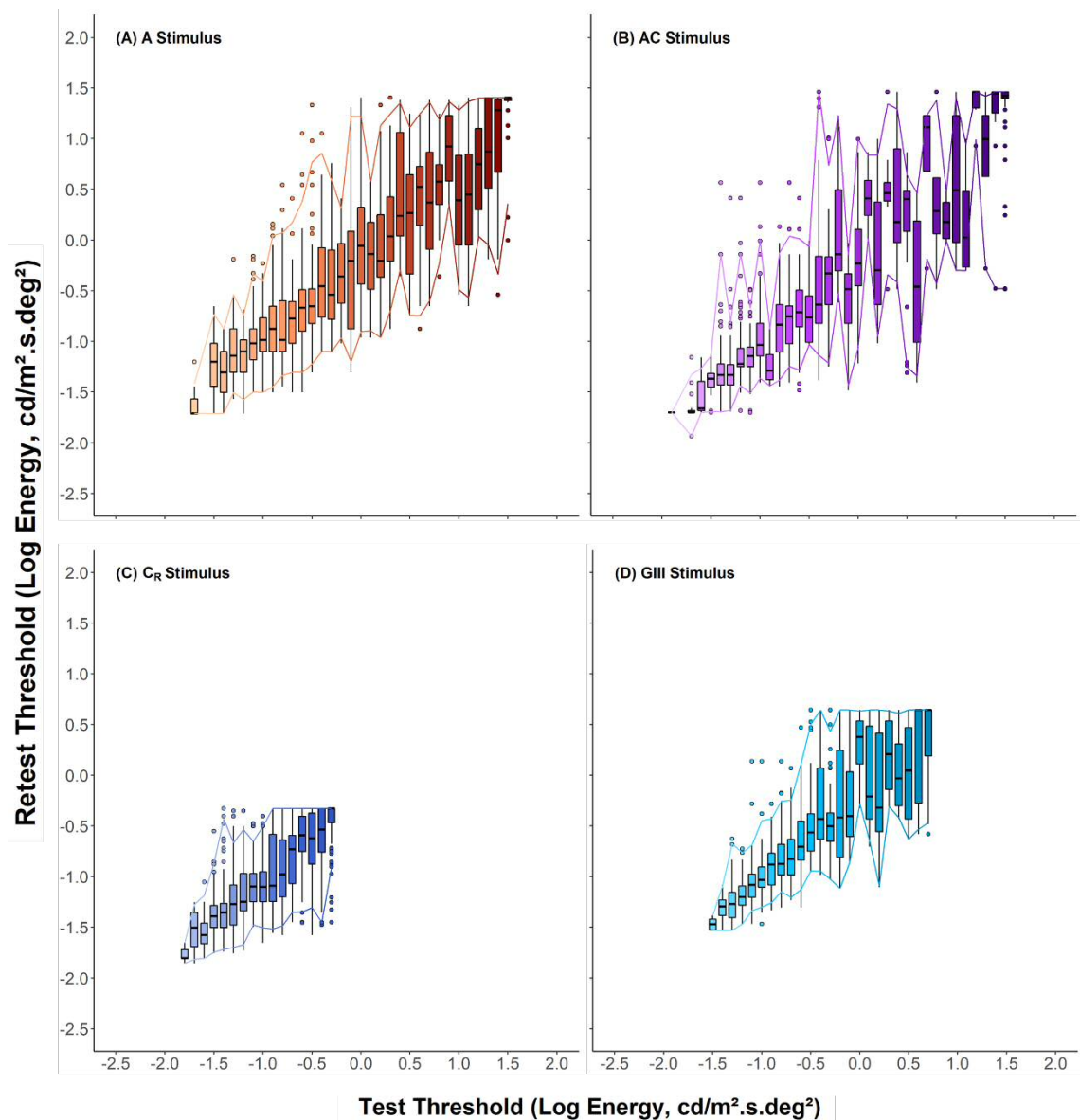
Figure 5.9 – Retest plotted against test thresholds (grouped in 0.1 log energy intervals) for the four locations tested in chapter four, (A) in Figure 5.1, for the four stimulus forms, incorporating thresholds from both glaucoma and healthy participants. Solid lines indicate the 5th and 95th retest percentiles.

The 5th and 95th retest percentiles

To enable accurate comparisons between stimuli, the 5th and 95th retest percentiles for each stimulus were plotted together (*Figure 5.10*). Although not directly applicable to the A, AC, and C_R stimuli, axes displaying the equivalent HFA dB for a Goldmann III stimulus are also shown.
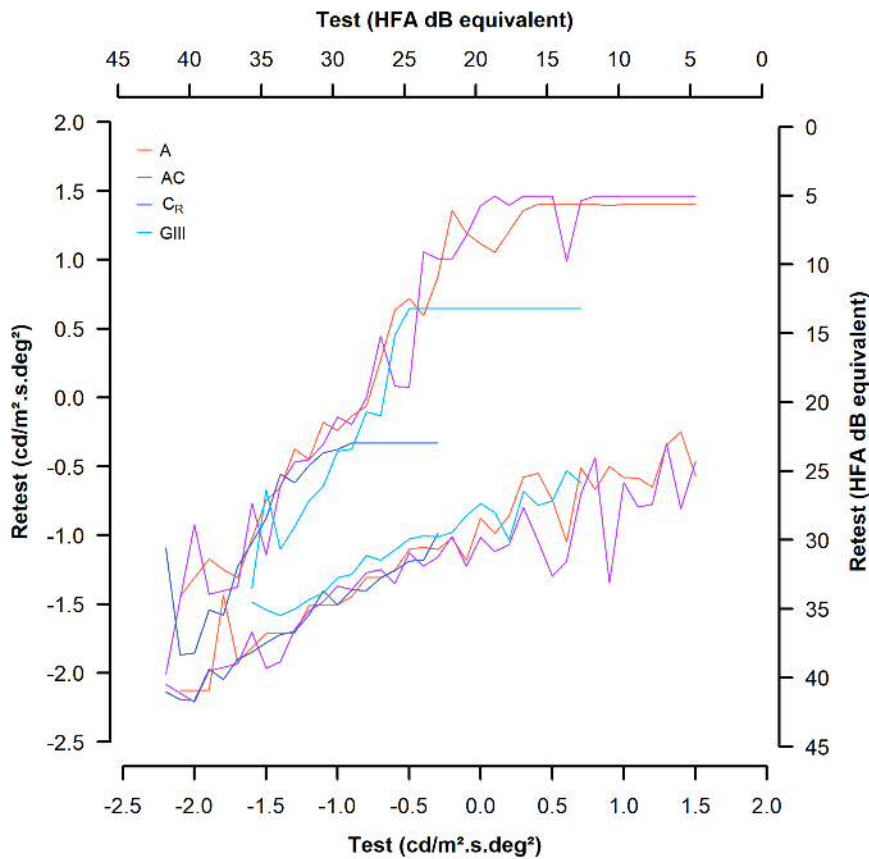
Figure 5.10 – 5th and 95th retest percentiles for each of the four stimulus forms.

As detailed in section *5.2.3.2*, the areas between the 5th and 95th retest percentiles were compared between the four stimulus forms. However, an accurate comparison could not be made from *Figure 5.10* for several reasons. First, the dynamic ranges of the four stimulus forms are quite different. *Figure 5.11* shows the 5th and 95th retest percentiles for each stimulus form, with their respective dynamic ranges indicated with dotted lines; as the A and AC stimuli have similar dynamic ranges in this experiment, only that for the A stimulus is indicated here. To accurately compare between the four stimulus forms, the maximum threshold for both test and retest were matched to that of the C_R stimulus, such that the areas between the 5th and 95th retest percentiles were determined only within the dark blue boundaries (the maximum log energy of the C_R stimulus) in *Figure 5.11*; this will be referred to as the 'C_R comparison'. To further investigate the A, AC and GIII stimuli, which had a greater dynamic range than that of

the C_R stimulus, comparisons were made between these stimuli within the light blue

boundaries (the maximum log energy of the GIII stimulus), referred to as the 'GIII

comparison', and comparisons were made between the A and AC stimuli within the

red boundaries (the maximum log energy of the A stimulus), referred to as the 'area

comparison'.
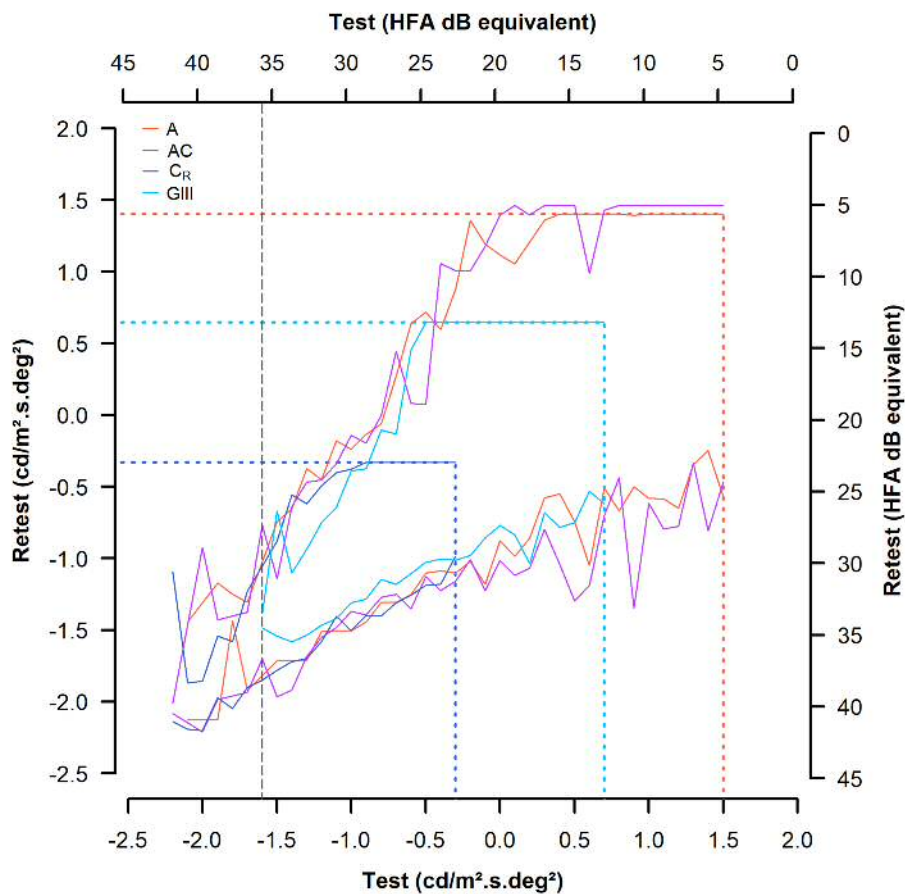
Figure 5.11 – 5th and 95th retest percentiles for each of the four stimulus forms. The coloured, dotted lines indicate the differing dynamic ranges for the stimulus forms. The grey, dashed line, indicates the minimum threshold determined with the GIII stimulus.

Second, the lowest threshold determined with the GIII stimulus was substantially

higher than that for the A, AC, and C_R stimuli, as indicated by the vertical, dashed, grey

line in *Figure 5.11*. As such, any comparison made as described above will determine a

smaller area between the 5th and 95th retest percentiles for the GIII stimulus, simply

due to the apparent 'extra' data to the left of this dashed line for the other three

stimuli. Consider the most sensitive test locations shown in *Figure 5.11*, i.e. those test locations with the lowest thresholds. With the GIII stimulus, the lowest measured threshold was -1.6 cd/m^2.s.deg^2, representing the thresholds for the most sensitive test locations with this stimulus. In comparison, the lowest measured threshold was -2.2 cd/m^2.s.deg^2 for the A, AC and C$_R$ stimuli; a test threshold of -1.6 cd/m^2.s.deg^2 represents a notably less sensitive test location with these three stimuli. Therefore, the same point on the graph does not represent the same test locations with the GIII stimulus, as with the other three stimuli. To overcome this problem, the 5th and 95th retest percentiles were transposed for the GIII stimulus, such that the lowest test threshold for this stimulus matched that for the other three stimulus forms (-2.2 cd/m^2.s.deg^2), ensuring that the same point on the graph represented the same test locations. Data were translated by equal amounts along the x- and y- axes. This is illustrated in *Figure 5.12*; the 5th and 95th retest percentiles for the C$_R$ stimulus are indicated by the dark blue solid lines, and the transposed 5th and 95th retest percentiles for the GIII stimulus are indicated by the light blue solid lines. The original 5th and 95th retest percentiles for the GIII stimulus are indicated by the light blue dashed lines. For ease of visualisation, the A and AC stimuli have not been included in this plot.
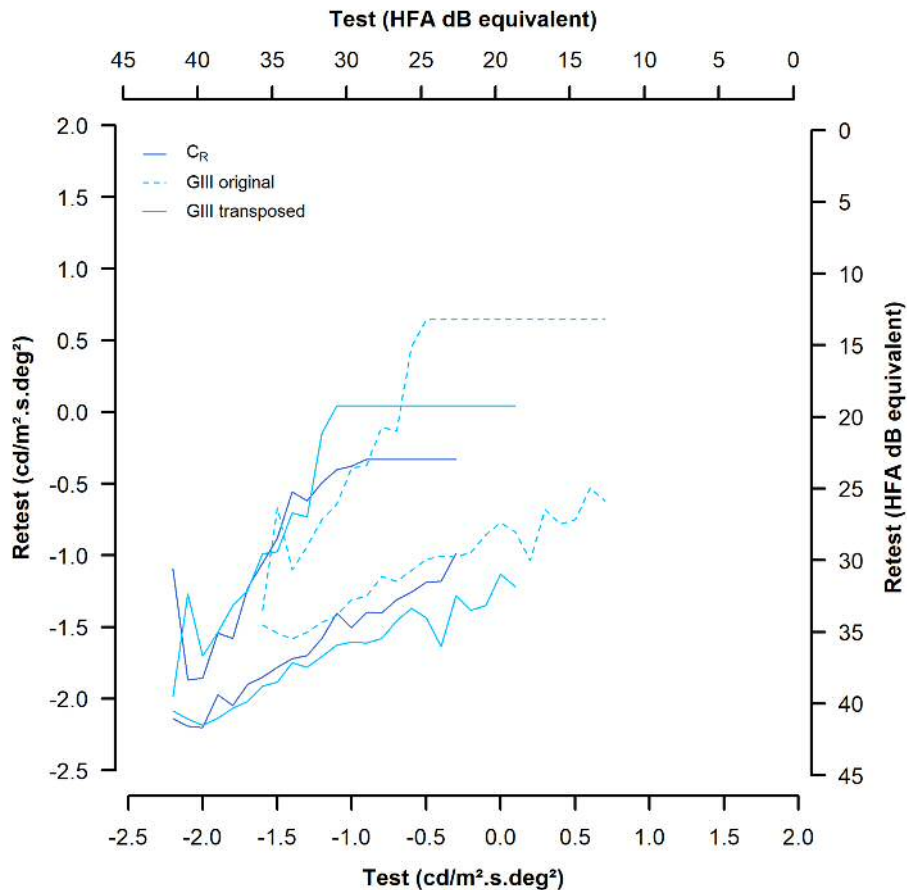
Figure 5.12 – 5th and 95th retest percentiles for the C_R and GIII stimulus forms. Values for the GIII stimulus have been transposed, such that the minimum test value for the 5th percentile matches that of the C_R stimulus. Original values are illustrated by the dashed lines.

Figure 5.13 illustrates the actions taken to overcome both the issues described. The 5th and 95th retest percentiles are shown for the four stimulus forms; those shown for the GIII stimulus are the transposed percentiles as described above. The areas between these percentiles were determined according to the shaded areas. For the C_R comparison, the areas between the 5th and 95th retest percentiles for all four stimulus forms were compared within the dark blue, shaded area, which represents the limits of the C_R stimulus. For the GIII comparison, the area between the 5th and 95th retest percentiles for the A, AC, and GIII stimulus forms were compared within both the dark blue, and the light blue, shaded areas, which represent the limits of the GIII stimulus. For the area comparison, the area between the 5th and 95th retest percentiles for the A, and AC stimulus forms were compared within all three shaded areas, which represent the limits of the A stimulus.
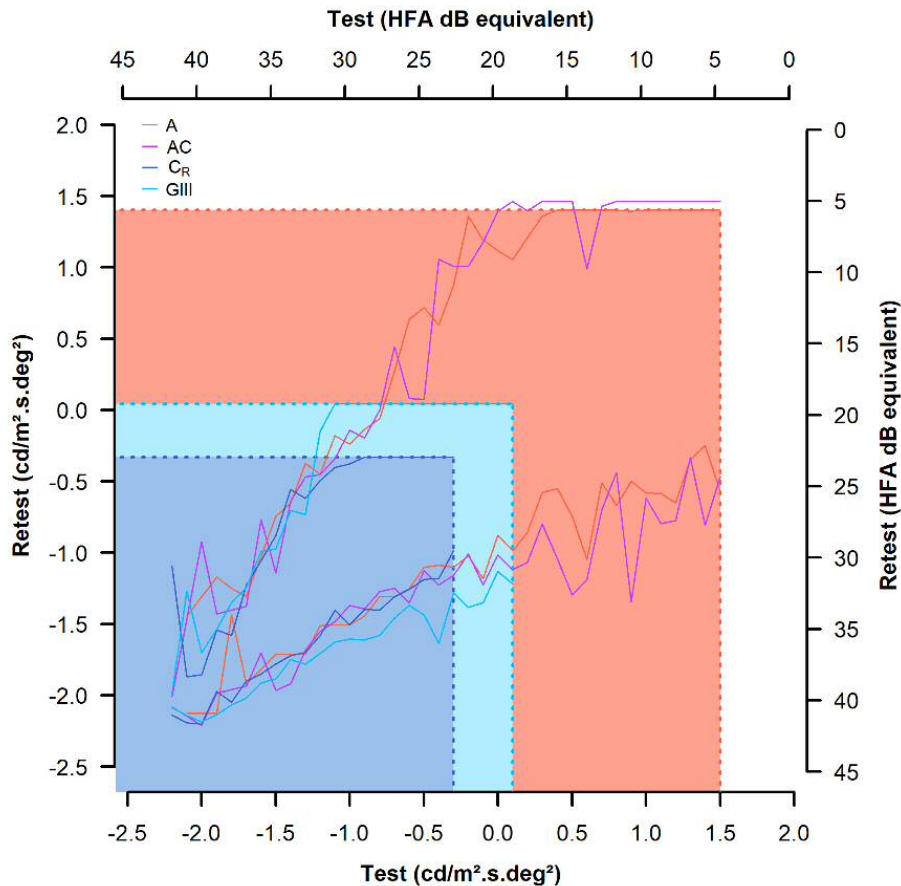
Figure 5.13 – 5th and 95th percentiles for each of the four stimulus forms; the transposed percentiles are shown for the GIII stimulus. The C_R comparison is indicated by the dark blue area, the GIII comparison is indicated by both the dark and light blue areas, and the area comparison is indicated by all three shaded areas.

Area between the 5th and 95th retest percentiles

The area between the 5th and 95th retest percentiles for each stimulus form was determined within each of the three comparisons (C_R comparison, GIII comparison and area comparison) as described, and are displayed in *Table 5.1*. From this table, it can be observed that the C_R stimulus had the smallest area between the 5th and 95th retest percentiles within the C_R comparison (1.60), the AC stimulus had the smallest area within the GIII comparison (2.42), and the A stimulus had the smallest area within the area comparison (5.79). The GIII stimulus had the greatest area within both the C_R (1.90) and GIII (2.76) comparisons.

Stimulus	C_R Comparison	GIII Comparison	Area Comparison
A	1.81	2.50	5.79*
AC	1.73	2.42*	5.86
GIII	1.90	2.76	
C_R	1.60*		

Table 5.1 – Area between the 5th and 95th retest percentiles for each of the four stimulus forms, within the three comparisons. The smallest area within each comparison is indicated by a *.

5.3.2.2 Complete and incomplete spatial summation

As described in section *5.2.3.2*, test locations were subdivided into two categories, those ≤ 31 dB, and those > 31 dB according to a modified perimetric sensitivity with the SITA Standard 24-2 program (HFA II). These two subcategories represented test locations which underwent complete (CSS), and incomplete (ISS) spatial summation respectively, with a Goldmann III stimulus at threshold. Given the method by which the A, AC and C_R stimuli were designed, as described in section *4.2.3*, it is expected that these stimulus forms undergo complete spatial summation at all test locations at threshold. Therefore, for those test locations categorised as undergoing complete spatial summation at threshold, this is true for all four stimulus forms, and for those test locations categorised as undergoing incomplete spatial summation at threshold, this is true for the GIII stimulus only.

A Goldmann III stimulus generally undergoes incomplete spatial summation within the central 15° of the visual field, and complete spatial summation at locations further from fixation, for a healthy observer (Swanson et al. 2004). Given the substantial number of test locations outside the central 15°, and/or damaged by glaucoma, there were a larger number of test locations included in the CSS category (347), compared with the ISS category (13).

As for the full dataset, all five permutations of test order were determined for both sub-categories of spatial summation, and the 5th and 95th retest percentiles for each stimulus form established (*Figure 5.14*). The area between these percentiles was then determined for the three comparisons, as for the full dataset (described above), and are displayed in *Table 5.2*.
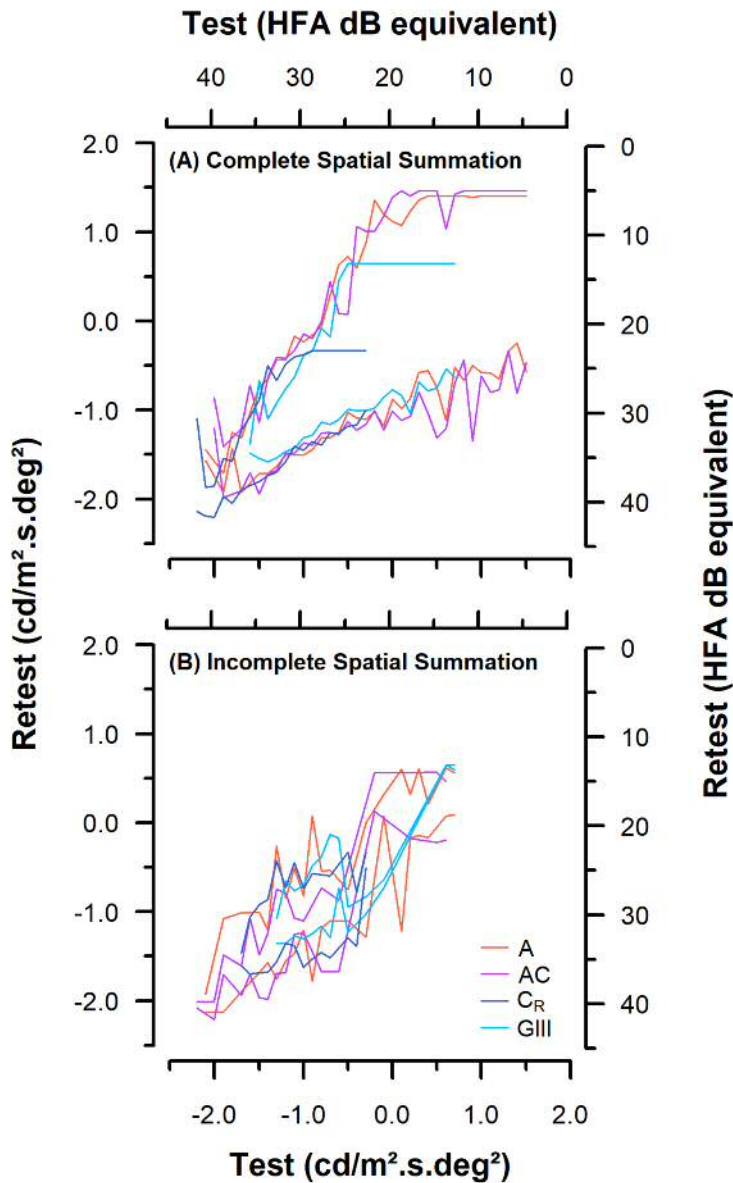
Figure 5.14 – 5th and 95th retest percentiles for each of the four stimulus forms for (A) CSS and (B) ISS test locations.

For CSS test locations, findings are similar to those for the complete dataset. From *Table 5.2*, it can be observed that, the C_R stimulus had the smallest area between the 5th and 95th retest percentiles within the C_R comparison (1.59), the AC stimulus had the smallest area within the GIII comparison (2.23), and the A stimulus had the smallest area within the area comparison (5.61). The GIII stimulus had the greatest area within both the C_R (1.89) and GIII (2.76) comparisons.

Findings differ for ISS test locations. From *Table 5.2*, it can be observed that the GIII stimulus had the smallest area between the 5th and 95th retest percentiles for the GIII comparison (0.65), and the AC stimulus had the smallest area between the 5th and 95th retest percentiles within the C_R comparison (0.50), and the area comparison (1.58).

Stimulus	C_R Comparison	GIII Comparison	Area Comparison
Complete Spatial Summation (CSS)			
A	1.62	2.31	5.69*
AC	1.62	2.23*	6.09
GIII	1.89	2.76	
C_R	1.59*		
Incomplete Spatial Summation (ISS)			
A	1.00	1.51	2.17
AC	0.50*	0.80	1.58*
GIII	0.62	0.65*	
C_R	1.12		

Table 5.2 – Area between the 5th and 95th retest percentiles for each of the four stimulus forms, for CSS and ISS test locations. The three comparisons are shown, and the smallest area within each comparison is indicated by a *.

5.4   Discussion

As detailed in section *5.1*, one of the main limitations of SAP is the high test-retest variability, which has been shown to systematically enlarge with decreasing perimetric sensitivity (Wilensky and Joondeph 1984; Heijl et al. 1989a; Artes et al. 2002a). Ideally, any new stimulus paradigm should address this limitation, demonstrating a lower test-retest variability than that of SAP, and remaining uniform across the range of thresholds.

In this experiment, the 5th and 95th retest percentiles were determined, and the area between these boundaries calculated for each of the four stimulus forms. These areas were compared between stimuli, such that the smallest area represented the lowest test-retest variability. As the dynamic ranges of the four stimulus forms were quite different, due to the hardware used in this experiment, three different comparisons were made between stimulus forms to ensure appropriate, and direct comparisons.

Areas between the 5th and 95th retest percentiles for all four stimulus forms were compared within the dynamic range limits of the C_R stimulus, referred to as the 'C_R comparison'. Beyond this threshold range, areas between the 5th and 95th retest percentiles for the GIII, A and AC stimuli were compared within the dynamic range limits of the GIII stimulus ('GIII comparison'), and for the A and AC stimuli were compared within the dynamic range limits of the A stimulus ('area comparison').

Examining data from all test locations, as shown in *Table 5.1*, the stimulus form identified as having the smallest area between the 5th and 95th retest percentiles differed in each of the three comparison categories. This same pattern was also observed when examining CSS test locations, which underwent complete spatial summation with a Goldmann III stimulus at threshold (*Table 5.2*). For ISS test locations, which underwent incomplete spatial summation with a Goldmann III stimulus, the observation is somewhat different; the AC stimulus had the smallest area within the C_R and area comparisons, and the GIII stimulus had the smallest area within the GIII comparison.

There does not appear to be one stimulus form that demonstrates the lowest test-retest variability in all comparisons. In fact, each stimulus form was determined as having the smallest area between the 5th and 95th retest percentiles in at least one comparison examined. Different stimulus forms were identified as having the smallest area between the 5th and 95th retest percentiles for CSS test locations in comparison with ISS test locations. This was true in all three comparisons conducted, however there was no particular trend identified that could attribute differing test-retest variability to spatial summation characteristics. It was originally hypothesised that stimuli optimised to probe spatial summation differences in the presence of glaucoma would demonstrate a lower test-retest variability, however this does not appear to be the case, as test-retest variability is largely similar between stimulus forms, and any differences observed are unlikely to be clinically significant.

Examination of CSS test locations identified the same stimuli as having the smallest area between the 5th and 95th retest percentiles as that for examination of all test locations together, in each of the three comparisons. This may be simply because

96.4% of the test locations were identified as undergoing complete spatial summation with a Goldmann III stimulus (347 test locations), while only 3.6% underwent incomplete spatial summation (13 test locations).

The hockey-stick model of Swanson et al. (2004), as shown in *Figure 5.3*, was found to fit 82% of the data they included in their study; while this is a high percentage, it is not a perfect fit. In addition, as described in section *5.2.3.2*, differing thresholding algorithms and test grids were used in the experiment presented here, compared with that of Swanson et al. (2004), and the GIII stimulus used in this experiment is slightly smaller than a Goldmann III stimulus. This being the case, and given that Ricco's area differs between individuals, there is likely to have been a misclassification of some test locations with respect to the type of spatial summation undergone with a Goldmann III stimulus and the GIII stimulus used in this experiment.

The test-retest variability characteristics observed in this experiment are similar to those noted by Artes et al. (2002a) with SITA Fast, SITA Standard, and Full Threshold testing strategies with a Goldmann III stimulus. As discussed in section *5.1*, each of these testing strategies establishes the perimetric sensitivity at four seed locations within the central 13°, then uses this information to determine appropriate stimulus presentations for adjacent test locations. Stimuli are therefore generally presented more centrally in the initial stages of the test, and more peripherally as the test proceeds, such that the observed increase in test-retest variability with eccentricity from fixation (Katz and Sommer 1986; Heijl et al. 1987; Heijl et al. 1989a; Heijl et al. 1989b; Chauhan and House 1991) could be explained by the fatigue effect with test duration. However, in this experiment, stimuli were presented at truly randomised test locations; as the same test-retest characteristics were still apparent, the reported increase in test-retest variability with eccentricity from fixation cannot solely be attributed to a fatigue effect.

The lowest measured threshold for the GIII stimulus was observed to be substantially higher than that for the other three stimulus forms (*Figure 5.10*); the lowest measured threshold represents those test locations with the highest sensitivity, i.e. near-normal test locations closer to fixation. These test locations will undergo complete spatial

summation with the A, AC, and C$_R$ stimuli, but incomplete spatial summation with the GIII stimulus. This difference in threshold is therefore not unexpected. For those stimuli undergoing complete spatial summation, luminance and area will be inversely proportional at threshold, such that threshold energy remains constant between stimuli. However, the GIII stimulus, which undergoes incomplete spatial summation at these locations, will demonstrate a higher threshold energy. This is illustrated in *Figure 5.15*, a schematic diagram which shows the spatial summation curve as log threshold energy plotted against log stimulus area.
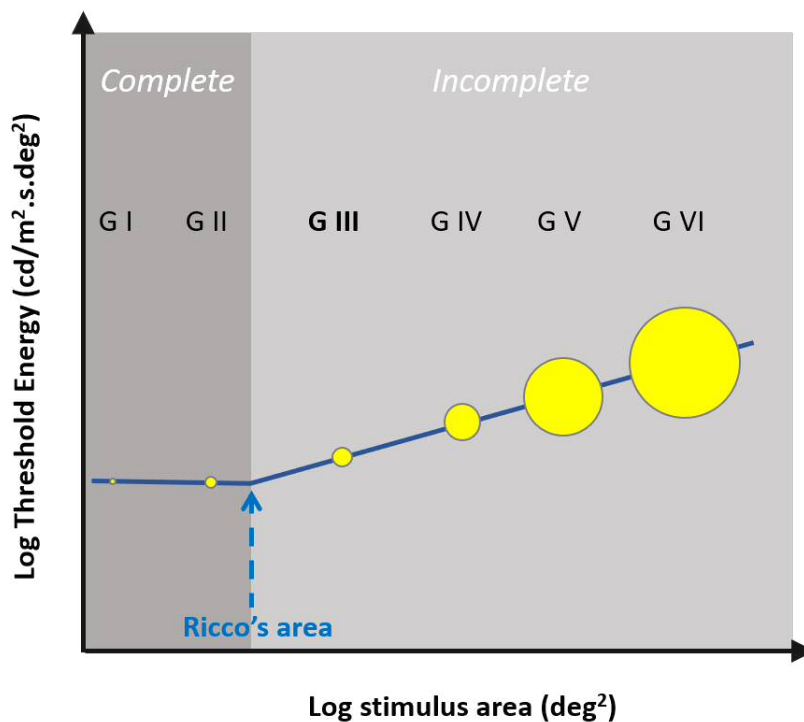
Figure 5.15 – A schematic diagram of a spatial summation curve, showing log threshold energy against log stimulus area.

However, as observed in *Figure 5.14*, this difference in lowest measured threshold is also observed in CSS test locations. If the above explanation held true, one would expect the lowest measured threshold to be approximately equal between all stimulus forms at CSS test locations, but the difference in lowest measured threshold would still be apparent at ISS test locations; this does not appear to be the case. As previously

stated, the GIII stimulus is slightly smaller than a standard Goldmann III stimulus, on which the CSS, or ISS classifications were made. It is therefore possible the GIII stimulus actually underwent complete spatial summation at some test locations classified as ISS. However, CSS test locations would still be classified correctly. It is possible this discrepancy between observed minimum threshold, and expected minimum threshold with the GIII stimulus may be explained by the potential misclassifications highlighted previously, resulting from differences between thresholding algorithms and test grids between this experiment and that of Swanson et al. (2004).

Findings for healthy participants indicated that there may be a greater uniformity in test-retest variability with threshold increase due to eccentricity from fixation (A and AC stimuli, *Figure 5.8.A* & *B*), although it was not possible to analyse this as outliers skewed the 5th and 95th retest percentiles. Findings for all participants at test locations equidistant from fixation indicated a higher test-retest variability with threshold increase due to glaucomatous damage, with all four stimulus forms (*Figure 5.9*). This may indicate that a greater test-retest variability observed with area-modulating stimuli is more likely to indicate cell dysfunction, although this would require further investigation before this could be accurately concluded. Given this observation, one would expect to observe a smaller area between the 5th and 95th retest percentiles for these stimuli, which was not found in all comparisons.

It is possible that the 5th and 95th retest percentiles demonstrated here are higher than the true retest values, given that participants completed four tests for the four different stimulus forms at each visit. Although randomising the order of these stimuli helps to ensure that neither learning nor fatigue effects unduly influence the results of one stimulus over another, it does not eliminate the fatigue effect itself, and the thresholds determined for the last test at a given visit may be higher than the thresholds determined for the first test simply due to the fatigue effect. In a clinical setting, patients would complete one test only for each eye at any one session; an experimental design that mimics this would be the most realistic method of examining test-retest limits. However, the purpose of this experiment was to compare test-retest variability between stimulus forms. Although the values themselves may be a little

higher than the true values, this will be the case for all stimulus forms, such that the comparisons presented here are still valid.

Comparison of mean thresholds across the 18 test locations using a one-way ANOVA with repeated measures, found no statistically significant difference between the five tests with any of the four stimulus forms, either for participants new to the A, AC, and C_R stimuli, or for participants with prior experience of these stimuli. The assumption of sphericity is particularly important in a one-way ANOVA with repeated measures, and in this experiment, it gave an indication as to whether fatigue effect may have had an inadvertent effect on test-retest variability. If a substantial difference in threshold occurs due to the test order (i.e. if threshold differences occur because a test for one stimulus is conducted first at one visit, but last at another visit), a violation of sphericity may be expected, as this difference is unlikely to be the same between participants, resulting in differing variances between test number. Mauchly's Test of Sphericity identified two occurrences of sphericity violation, the AC stimulus in experienced participants, and the C_R stimulus in novice participants. One might expect that, if fatigue effect caused a substantial difference in threshold measurements in this experiment, Mauchly's Test of Sphericity would be violated in all four stimulus forms, and in both groups of participants. Equally, if one stimulus was more affected by fatigue effect than others, one would expect a consistent stimulus to be identified as violating sphericity in both groups of participants, which was not observed here. It is therefore unlikely that fatigue effect has substantially increased the test-retest variability observed in this experiment.

It is acknowledged that Mauchly's Test of Sphericity is not without limitation itself. However, as the risk associated with sphericity violation is an increase in type I error (i.e. a mistaken finding of statistical significance), any violation of sphericity not correctly identified by Mauchly's Test of Sphericity in this experiment has not resulted in any adverse findings in the investigation of learning/fatigue effect.

The lack of statistical significance with a one-way ANOVA with repeated measures was interpreted as there being no significant learning or fatigue effect with any of the four stimulus forms, although it is unclear whether this translates as an absence of

substantial learning/fatigue effect at all, or whether prior experience with SAP is sufficient to negate any learning effect with these stimuli. Given these findings, one could hypothesise that the same experiment, conducted with participants who were completely new to perimetry, should yield similar results to those observed in prior studies conducted into learning effects with SAP, in that a minority of observers would display a substantial learning effect, but that the majority would produce reliable results with little practise (Heijl et al. 1989b). This experiment should be conducted before drawing this conclusion, although it is reassuring that no re-learning would be required for those observers with prior SAP experience, should the A stimulus (with which a greater SNR was demonstrated in chapter four) be deemed of sufficient benefit to warrant clinical introduction.

A study by Gardiner et al. (2014) argued that perimetric sensitivities less than ~19 dB could not be reliably determined due to saturation of the retinal ganglion cells by high stimulus luminance. They hypothesised that, once the ganglion cell reaches saturation, further increase of the stimulus luminance would have no effect, and no further cell response would be elicited. They therefore determined that this was the reason for the increase in test-retest variability at locations of low perimetric sensitivity. A more recent study by Anderson et al. (2016), examining healthy observers, determined that no such saturation was observed in the healthy retina, but did not test those with glaucomatous defects; although this study challenged the argument of Gardiner et al. (2014), the possibility of saturation in dysfunctional retinal ganglion cells could not be ruled out completely.

The findings of this experiment, which demonstrate similar test-retest characteristics with all four stimulus forms, support the findings of Anderson et al. (2016), and provide further evidence that ganglion cell saturation is not the cause of a greater test-retest variability with decreasing perimetric sensitivity. The A stimulus presented here, in which luminance remained constant throughout the test, showed the same increase in test-retest variability with increasing threshold (decreasing sensitivity) as the other three stimulus forms, in which luminance modulated to establish threshold. If retinal ganglion cell saturation occurred with high stimulus luminance, a more uniform test-

retest variability would have been demonstrable with the A stimulus at those locations with a higher threshold, which was not observed.

There are some limitations in this experiment that should be noted. In using a flat OLED display to present stimuli, these stimuli will be subject to some distortions of shape and luminance due to viewing angle. As reported by Ito et al. (2013), luminance from an OLED display decreases as the viewing angle increases. Due to limitations of the apparatus available, differences in luminance with viewing angle could not accurately be measured with the Sony PVM-A250 Trimaster El OLED display used in this series of experiments. In addition, a circular stimulus presented in the periphery on a flat display will create an elliptical, not circular, image on the retina. In contrast to the experiment presented in chapter four, in which the four test locations were at 9.9°, equidistant from fixation, thus largely controlling for these distortions, this experiment presented stimuli at a range of test locations and eccentricities. No correction has been applied to measured threshold values to account for these distortions. However, this experiment investigated the variability of threshold measurements between five tests; the raw threshold measurements themselves were not the focus, but the variability in threshold measurements from one test to another. Given that test locations remained constant, and all participants completed tests for each stimulus form at each visit, thus serving as their own control, it is not expected that findings were unduly influenced.

The results of this experiment indicate no clinically significant difference in test-retest variability characteristics due to the stimulus, for the four stimulus forms investigated here. Given that the A stimulus has shown a greater SNR than that of a Goldmann III equivalent stimulus (chapter four), further investigation is recommended with the application of a clinical thresholding algorithm, which would permit thresholds for more locations to be determined within an acceptable test-time. As existing thresholding algorithms were originally developed for use with a Goldmann III stimulus, a newly designed thresholding algorithm is likely necessary for use with the A stimulus, the development of which is outside the scope of this study. Once a suitable thresholding algorithm has been determined, it would be advisable to repeat this

experiment, to establish the expected test-retest variability characteristics in a clinical setting.

Chapter 6 Resistance of perimetric stimuli, optimised to probe changing spatial summation in glaucoma, to optical defocus and intraocular straylight

6.1   Introduction

When examining the visual field to identify possible change, it is important to successfully distinguish neural visual loss, such as occurs in glaucoma, from the effects of optical imperfections, such as may occur in the presence of cataract, uncorrected refractive error, or corneal opacities. Any perimetric test aimed at identifying neural damage should be as robust as possible to these optical imperfections, whilst remaining sensitive to early, subtle neural changes, however there are many studies which highlight the difficulties of this task. Here, two causes of optical imperfection are investigated, optical defocus and intraocular straylight, which have been shown to impact upon perimetric results in previous studies.

6.1.1   Optical defocus

Optical defocus can be caused by both over- and under- correction of refractive error in presbyopic patients in whom there is little or no remaining accommodation; this is the typical patient demographic undergoing perimetric tests for glaucoma.

The impact of blur on perimetric results is well documented. Studies generally agree that smaller stimuli, which display higher spatial frequency characteristics, are more vulnerable to defocus than larger stimuli, which display lower spatial frequency characteristics (Sloan 1961; Campbell and Green 1965; Atchison 1987; Anderson et al. 2001; Horner et al. 2013), and that the effect of defocus is more pronounced at locations closer to fixation (Sloan 1961; Fankhauser and Enoch 1962; Benedetto and Cyrlin 1985; Atchison 1987; Anderson et al. 2001). SWAP has been reported as resistant to blur up to +3.00 DS, likely due to the large stimulus area used (Herse et al. 1998), but this was only measured in the central 10°.

Contrary to those studies listed above, which reported a more pronounced effect of defocus at locations closer to fixation, Henson and Morris (1993) found that threshold elevation with dioptres of defocus was independent of eccentricity, although this may

be due to the lower background luminance used (0.25 cd/m^2) compared with other studies (10 cd/m^2).

There is some discrepancy between studies with respect to the stimulus area at which optical blur no longer causes a statistically significant threshold difference, and how many dioptres of defocus that stimulus is robust to. Sloan (1961) did not find a statistically significant effect on threshold with stimuli ≥ 0.15 deg^2 (Goldmann III), even at the fovea, up to 3.00DS of optical blur, while Atchison (1987) reported a statistically significant difference in threshold with a Goldmann III stimulus at equivalent levels of blur. Benedetto and Cyril (1985) reported that 3.00 DS of blur was required to cause a statistically significant difference in threshold within the central 12° (Goldmann III), whereas Heuer et al. (1987) found that only 1.00 DS of defocus was required for the threshold difference to be statistically significant with the same stimulus area. In contrast to Sloan (1961) and Atchison (1987), Heuer et al. (1987) found that locations closer to fixation appeared more robust to the effects of optical defocus. It is likely that these discrepancies are due in part to slight differences in study design. For example, Sloan (1961) tested only one participant, and only along the horizontal meridian, and Atchison (1987) tested five participants along the vertical meridian, with one participant undergoing additional tests along the horizontal meridian. Heuer et al. (1987) tested five participants along the horizontal meridian in the nasal visual field only, and was the only one of these three studies to use cycloplegia. It is also likely that discrepancies between studies were in part due to inter-observer differences, which Atchison (1987) reported to be statistically significant.

Confounding factors have also been investigated to determine their impact, and distinguish their effects from that of optical defocus. These include pupil diameter, and peripheral refractive error, which has been found to differ from foveal refractive error (Ferree et al. 1931; Ferree and Rand 1933; Rempt et al. 1971; Millodot 1981), in addition to magnification/minification and prismatic effects from the use of lenses both to correct refractive error, and to induce blur. Henson and Morris (1993) and Herse (1992) reported that the difference in threshold with optical defocus was dependent upon pupil diameter. Brenton and Phelps (1986) reported that pupil diameter did not influence the mean sensitivity of the visual field, although noted that

pupil diameter was statistically significantly related to age, which in turn was statistically significantly related to perimetric sensitivity. Anderson et al. (2001) investigated the effects of establishing and correcting peripheral refractive error for a test location at 30° eccentricity, and found their results to be in general agreement to those studies in which peripheral refractive error had not been considered. Atchison (1987) conducted his study using contact lenses to induce blur, thus eliminating magnification and minification effects, and reducing peripheral aberrations as induced by spectacle lenses, and generally found his results to agree with those of other studies that utilised trial lenses. With respect to prismatic effect, Anderson et al. (2001) positioned the trial lenses as to eliminate prismatic effects, and found similar results to those reported elsewhere. Atchison (1980) also found the influence of prismatic effects from trial lenses to be relatively small in comparison to the effects of defocus, in those lenses < ± 10.00 DS, concluding that these effects could be largely ignored in static perimetry.

Grating stimuli, such as those utilised in FDT, have been reported as robust to the effects of optical blur at the fovea (Anderson and Johnson 2003), whereas Artes et al. (2003b), examining a greater number of test locations out to an eccentricity of 30°, found a small reduction in perimetric sensitivity with optical defocus. Horner et al. (2013) reported that grating stimuli were robust to peripheral defocus, although they found this to be highly dependent upon the spatial frequency of the stimulus. The task in this study was one of grating contrast detection.

Peripheral resolution acuity for high contrast gratings is reportedly substantially lower than central acuity, even with correction of peripheral refractive error, as the sampling density of the underlying retinal photoreceptors and ganglion cells declines more quickly than optical quality outside the fovea. Spatial frequencies higher than the neural sampling limit (Nyquist frequency) of the retina thus result in observations of aliasing, where the under-sampled high spatial frequencies appear as lower spatial frequencies and of non-veridical orientation (Thibos et al. 1987; Anderson and Hess 1990; Anderson 1996; Thibos et al. 1996). The sampling- rather than optically- limited nature of peripheral resolution may explain why the effects of optical defocus are less pronounced for peripherally presented grating or letter stimuli. Anderson (1996) and

Wang et al. (1997) additionally noted that, although detection sensitivity was statistically significantly higher than resolution sensitivity for grating and letter stimuli up to 40° eccentricity, resolution acuity was less affected by optical blur, again likely because the limiting factor for resolution is not contrast but the underlying sampling density of the retina. None of these studies compared their results directly to a standard Goldmann III stimulus.

It may be concluded that optical defocus does have some effect on perimetric findings, although there appears to be some variation between studies as to the exact effect, possibly due to inter-individual variability, the differences between study designs noted above, and the differing robustness of different stimuli and tasks to defocus. It is important to quantify these effects, particularly for a novel stimulus in which this has not previously been investigated, given that optical defocus is largely unavoidable in clinical practice. Perimetry in some instances may be conducted without appropriate correction of refractive error, e.g. if a patient is tested in a hospital setting without current refractive error information, or if perimetry is conducted in the initial stages of an eye examination prior to establishing refractive error status. Human error may result in the use of an incorrect trial lens, e.g. selecting a negative instead of a positive trial lens, or the choice of full aperture trial lenses suitable for use with a perimeter may be somewhat limited.

Even if practitioners are careful to fully correct the refractive error for the appropriate working distance, this only corrects the foveal refractive error, and does not account for the differing refractive errors in the periphery, which may vary substantially from one location to another (Ferree et al. 1931; Ferree and Rand 1933; Rempt et al. 1971; Millodot 1981). These studies have reported peripheral refraction to vary from central measurements in three distinct ways: (i) an increase in myopia with increased eccentricity in the horizontal meridian, and an increase in hyperopia with increased eccentricity in the vertical meridian, (ii) an increase in hyperopia in both horizontal and vertical meridians, and (iii) asymmetry between the nasal and temporal meridians. The type of peripheral refractive error was noted to relate to central refractive error characteristics. The type of peripheral refraction was attributed to the combination of refractive characteristics of the cornea and the lens (which may vary with

accommodation), and the axial length and shape of the globe, although studies noted that the amount of refractive error difference with eccentricity was not readily predictable from central refractive error.

Atchison et al. (2006) noted a general myopic shift in the periphery in eyes that were centrally emmetropic, and a hyperopic shift in the periphery of those eyes that were centrally myopic, although, as in the studies noted above, the amount of shift was more difficult to predict from central refractive error. They also noted that myopia had a greater effect on peripheral refraction along the horizontal, compared with the vertical meridian, and that astigmatism was notably asymmetric between nasal and temporal fields.

Charman and Jennings (2006) investigated the longitudinal changes in peripheral refraction over a 25-year period, and noted relatively small changes over time, which they attributed to aging corneal and lenticular changes. They also noted that these changes in peripheral refraction did not appear to account for the reduction in peripheral sensitivity observed with age.

Charman and Radhakrishnan (2010) observed that one of the major difficulties in studies examining peripheral refractive error is that, whether objective or subjective, most of the instruments used were not designed for this purpose. This may, in part, explain the high inter-subject variability noted in other studies (Atchison et al. 2006).

It is not possible to correct refractive error at all locations simultaneously when conducting perimetric tests. This is partially due to the difficulty in predicting peripheral refractive errors from the central refractive error, partially due to the variation and asymmetry in peripheral refractive errors corresponding to different test locations, and, perhaps most crucially, due to the present lack of an optical system capable of correcting such differing refractive errors simultaneously. Perimetric tests therefore always involve sensitivity measurements of at least some test locations subject to optical defocus.

6.1.2 Straylight

The term 'straylight' here describes forward intraocular light scatter that results in the dispersion of light rays entering the eye to other areas of the retina, such that they are not involved in the formation of the normal image (Van den Berg 1986). Straylight at the retina may originate from a variety of sources; it has been found to increase in normal eyes with age (Ijspeert et al. 1990), and is greater with lower levels of ocular pigmentation (Van den Berg et al. 1991), as well as in the presence of media imperfections, such as cataract (Van den Berg 1986; De Waard et al. 1992), corneal dystrophies (Van den Berg 1986), corneal oedema (Fonn et al. 1999), and posterior capsule opacification (Meacock et al. 2003).

Many studies have investigated the effects of cataract on perimetric sensitivities. SAP sensitivities are known to be affected by the presence of cataract, resulting in a general reduction in perimetric sensitivity values, attributed to a reduction in luminance contrast between stimulus and background (Lam et al. 1991; Moss et al. 1995). Pattern deviation maps and the global pattern standard deviation index are often used as an attempt to distinguish glaucomatous field loss from that induced by cataract, or indeed optical defocus; these attempt to correct for a generalised depression in the visual field, highlighting only focal loss (Bengtsson et al. 1997b). However, the success of this approach in separating neural from optical visual loss is not clear. It is generally reported that pattern standard deviation, and pattern deviation maps remain largely unchanged following cataract surgery, whereas mean deviation and total deviation maps tend to show an improvement in sensitivity (Lam et al. 1991; Bengtsson et al. 1997b; Kim et al. 2001; Kook et al. 2004). In contrast, while Smith et al. (1997) and Hayashi et al. (2001) also noted an improvement in mean deviation, both studies reported that mean and corrected pattern standard deviation worsened following cataract surgery, concluding that cataract may mask true progression of visual field loss in glaucoma, despite the use of these algorithms. In addition, studies have found that almost all glaucomatous visual field progression comprised both focal and diffuse components (Henson et al. 1999; Artes et al. 2005a; Artes et al. 2010), indicating that reliance on pattern deviation analysis alone may overlook diffuse glaucomatous loss. Given such issues, neural visual loss may be

difficult to accurately differentiate from optical visual loss resulting from straylight, with conventional SAP.

A few studies have evaluated the effect of cataract on perimetric sensitivities obtained with stimuli of differing area. In a study of Goldmann kinetic perimetry, Radius et al. (1978) noted that smaller Goldmann stimuli were sub-threshold in the presence of cataract, whereas larger Goldmann stimuli were detected. Wood et al. (1989) tested participants with asymmetric amounts of lenticular opacity between the two eyes. They used two perimetric instruments, the Octopus 201, which utilised a standard Goldmann III stimulus of 0.43° diameter, and the Dicon AP3000, which utilised a stimulus diameter of 0.28°. Comparing the eye with the denser lenticular opacity against that of the eye with the less dense opacity, they noted a reduced perimetric sensitivity with increase in lenticular opacity with both stimuli, although noted differing effects with eccentricity. For the stimulus of diameter 0.43°, a greater reduction in perimetric sensitivity was noted with increase in eccentricity for observers with non-nuclear lenticular opacities, but noted that reduction in perimetric sensitivity was greatest centrally, decreasing with eccentricity, for observers with nuclear lenticular opacities. In contrast, for the stimulus of diameter 0.28°, the reduction in perimetric sensitivity was found to be greatest centrally, decreasing with eccentricity, for observers with both nuclear and non-nuclear lenticular opacities. It should be remembered however, that there are further differences between these two perimetric instruments than simply stimulus area, and neither instrument adheres to the standard background luminance (10 cd/m^2) or stimulus duration (0.2 seconds) of current SAP. The Octopus 201 uses a projection system to display stimuli, with a background luminance of 1.3 cd/m^2, and a stimulus duration of 0.1 seconds. In contrast, the Dicon AP3000 employs LED stimuli, with a stimulus duration of 0.4 seconds; two different background luminances of 3.2 cd/m^2, and 14.3 cd/m^2, were tested in this study. As such, it is difficult to ascertain whether the differences noted in this study are due to the differing stimulus areas, or to the confounding factors noted here.

Intraocular straylight produces a veiling luminance, reducing the contrast of the retinal image (De Waard et al. 1992). Various studies have compared contrast sensitivity

functions, determining contrast sensitivity for a range of gratings from low to high spatial frequency, in both eyes in participants with uniocular cataracts, and between participants with cataract and age-similar healthy participants (Hess and Woo 1978; Elliott et al. 1989; Pardhan and Gilchrist 1991; Drews-Bankiewicz et al. 1992; Elliott 1993). In most cases, a greater reduction in contrast sensitivity was found for medium and higher spatial frequencies in the presence of cataract than for lower spatial frequencies, although a uniform reduction in contrast sensitivity was found for all spatial frequencies in some participants. It was speculated that this more uniform reduction in contrast sensitivity, independent of spatial frequency, may be indicative of capsular involvement (Elliott et al. 1989).

The point spread function is known to widen and flatten in the presence of intraocular straylight (Van den Berg 1995; Bergin et al. 2011; Van den Berg 2017). Applying this to perimetric stimuli, the result would be a wider and lower-contrast representation of the nominal stimulus area and contrast on the retina. The percentage change in the stimulus area and contrast will be greater for a smaller stimulus than for a large one. Considering the effect of introducing a straylight source when viewing two stimulus (one small and one large) that are at their respective thresholds, it would be expected that, to restore threshold, a greater compensatory increase in contrast would be required for the small stimulus than for the large one, due to normal spatial summation. This differential effect would be observed only when the retinal representation of one of the filtered stimuli undergoes complete spatial summation (i.e. when it is smaller than Ricco's area). This may explain the finding of a greater reduction in contrast sensitivity observed for higher spatial frequencies in the presence of cataract than for lower spatial frequencies in some studies, but also the absence of such a finding in other studies.

Alternative perimetric stimuli, such as those utilised in SWAP and FDT, have been reported as more vulnerable to the effects of cataract. Kim et al. (2001) noted a greater improvement in mean deviation following cataract surgery in SWAP compared with SAP in healthy participants, by a factor of 2.4, indicating that SWAP sensitivity values were more affected by the presence of lenticular opacities than those of SAP. This is generally thought to be due to differences in the absorption characteristics of

the intraocular lens, due to both age and the presence of lenticular opacities. This was corroborated by Moss et al. (1995) who corrected their blue-on-yellow sensitivity values for ocular media absorption, and observed similar mean deviation with both blue-on-yellow, and white-on-white stimuli.

Several studies have conducted FDT in otherwise healthy participants before and after cataract extraction, and have noted a statistically significant improvement in mean deviation following cataract surgery, but no such difference in pattern standard deviation (Kook et al. 2004; Tanna et al. 2004; Ueda et al. 2006). Siddiqui et al. (2005) conducted FDT before and after cataract extraction in participants with glaucoma; they also noted a statistically significant improvement in mean deviation following cataract surgery, but in contrast to those studies conducted with healthy participants, they found a statistically significant decline in pattern standard deviation. As previously noted with SAP, this finding may indicate a masking of true progression of visual field loss due to glaucoma in the presence of cataract. Casson and James (2006) noted that posterior subcapsular lens opacities had a greater effect on FDT, resulting in a higher prevalence of visual field defects (which were then absent post-cataract extraction), in comparison with nuclear sclerotic and cortical lens opacities.

As discussed by Budenz et al. (1993), lens opacifications are thought to degrade the retinal image by three methods: (i) increased intraocular straylight, (ii) image blur, and (iii) reduction in illumination due to increased media absorption. The greatest effect is due to veiling glare from intraocular straylight, but it is not possible to separate these effects in those with existing lenticular opacities. As such, a comparison of perimetric results pre- and post- cataract surgery, as described in the studies above, is an imperfect method of analysing the differing effects of cataract on perimetry, as it is not clear from these studies which aspects of retinal image degradation are being investigated. Several studies (Heur et al. 1988; Budenz et al. 1993; Anderson et al. 2009; Bergin et al. 2001) have attempted to isolate the effects of straylight on perimetry by using various types of lenses to simulate these effects. This method has the added benefit of permitting the investigation of the effects on perimetric findings of more subtle, age-related changes in the optical quality of the intraocular lens, in

addition to the effects of intraocular straylight from more clinically significant lenticular opacification.

A number of studies have used ground-glass diffusing lenses, which Heur et al. (1988) confirmed increase forward light scatter with little effect on the size and shape of the image, illumination, or visual acuity; they noted a halo of scattered light surrounding the image, but it was > 10 times dimmer than the image itself. Heuer et al. (1988) used these lenses to investigate the effects of intraocular straylight on SAP sensitivities in healthy participants, and found that all diffusing lenses used resulted in a statistically significant reduction in perimetric sensitivity at all eccentricities tested between 0° and 25°. Budenz et al. (1993) used one of these diffusing lenses to investigate the effects of intraocular straylight on SAP sensitivities in those with glaucoma, and found a statistically significant decline in mean deviation, but no such effect on pattern standard deviation, such that normal areas of the visual field appeared equally affected by the presence of the diffusing lens as those areas with a glaucomatous defect.

Ground glass diffusing lenses are not a particularly good simulation of cataract as they do not usually display the wide-angle scatter, caused by opaque, light scattering 'discontinuities', typically observed with cataract. Anderson et al. (2009) and Bergin et al. (2011) used five white filters that contained light-scattering particles within the lens itself, rather than etchings on the surface as with ground-glass lenses. These lenses result in a wide-angle scatter. While Heuer et al. (1988) and Budenz et al. (1993) established the increased straylight with each diffusing lens using an optical bench, Anderson et al. (2009) and Bergin et al. (2011) used a commercially available straylight meter (C-Quant; Oculus, Wetzlar, Germany) to quantify the effects of induced straylight with each filter within the eye itself; a description of this instrument is provided in section *6.4.1.1*. They then investigated the effects of this straylight on several perimetric test strategies. Anderson et al. (2009) noted a greater reduction in perimetric sensitivity with SWAP, and particularly FDT, in comparison with SAP. In comparing perimetric sensitivity with each filter to that with no filter in place (i.e. baseline), FDT sensitivity was found to be statistically significantly reduced from baseline with the first, least dense filter. SWAP sensitivity was statistically significantly

reduced from baseline with the second filter, and SAP sensitivity was statistically significantly reduced from baseline with the third filter. Bergin et al. (2011) noted that mean deviation was statistically significantly reduced in FDT and SAP with increasing straylight, and also noted statistically significant differences in the effects of straylight between central and peripheral test locations, although have not defined the eccentricity limits of these two categories. These two studies also investigated more novel methods of perimetry, finding that Grating Resolution Perimetry (GRP), and the Moorfields Motion Displacement Test (MDT) were more robust to induced straylight.

However, Anderson et al. (2009) acknowledged that a comparison of differing perimetric strategies such as these is not a straightforward process, and noted that, if perimetric sensitivities were compared within the 95% confidence limits for sensitivity, only a minority of data-points fell outside this range.

As perimetric tests are primarily used in the older population, higher levels of straylight, either due to normal aging changes in the intraocular lens, or due to cataract, are to be expected. As such, a test which is less robust to intraocular straylight than the current standard may be of limited value, and it is important to quantify the effects of intraocular straylight in any new perimetric paradigm.

6.2 Experiments

Two experiments were undertaken to investigate the effects of optical defocus, and intraocular straylight, on the same four stimulus forms (A, AC, C_R and GIII), used in the experiments presented in chapters four and five. Given that larger stimuli have been reported as more robust to the effects of optical defocus (Sloan 1961; Campbell and Green 1965; Atchison 1987; Anderson et al. 2001; Horner et al. 2013), and there are some indications that larger stimuli may also be more robust to intraocular straylight (Radius 1978), it was hypothesised that the GIII stimulus would be more robust to these optical imperfections than the C_R stimulus.

It was more difficult to hypothesise how these optical imperfections may affect the area-modulating stimuli. If the effects of optical imperfections are primarily due to stimulus area, it may be that these effects will differ with threshold. Peripheral

locations, at which threshold will be higher (i.e. area-modulating stimuli will be larger at threshold), may be more robust to optical imperfections than central locations. Additionally, with an area-modulating paradigm, a curvilinear effect on threshold with increase in optical imperfections may be observed, whereby an initial increase in optical imperfections would cause a greater increase in threshold, with this increase in threshold becoming less pronounced with further increase in optical imperfections.

6.2.1　Overall methods

6.2.1.1 Apparatus and set-up

As detailed in chapter two, stimuli were presented on a gamma-corrected, 25" OLED display (Sony PVM-A250 Trimaster El, resolution 1920 x 1080 pixels, frame rate 60 Hz, refresh rate 120 Hz), driven by a ViSaGe MKII Stimulus Generator (Cambridge Research Systems, Rochester, UK). Experiments were programmed in MATLAB (version 2014b; The Mathworks, Inc., Natick, MA) using the CRS toolbox (version 1.27, Cambridge Research Systems, Rochester, UK), adapted from code supplied by T. Redmond and P. Mulholland. A nominally uniform background luminance of 10 cd/m^2 was used. During all tests, participants were instructed to fixate a central cross on the screen and respond to any stimulus they had detected in their visual field by pressing a button on a response pad (Cedrus RB-530; Cedrus, USA). Stimulus duration was fixed at 0.2 seconds in all experiments.

Details of testing protocols for the two experiments are given in sections *6.3.1* and *6.4.1*.

6.2.1.2 Participants

Both experiments were conducted with young, healthy participants, who had not been diagnosed with any systemic conditions, nor were taking any medications. Ocular health was confirmed via slit lamp biomicroscopy at each visit; no ocular health concerns were noted, and no corneal/lens opacities were observed by this author. No participant had undergone any ocular surgery, or had any first-degree relatives with glaucoma. IOP was confirmed as less than 21 mmHg at each visit.

Only one eye of each participant was tested; this eye was selected as the eye that best met the inclusion/exclusion criteria, or was selected at random if both eyes were

equally suitable. A full subjective refraction was conducted prior to the commencement of any experimental tests, and all participants had a best-corrected visual acuity of 6/6 or better. All participants had a normal visual field ('within normal limits') with the SITA Standard 24-2 program on the Humphrey Field Analyzer (HFA II, Carl Zeiss Meditec Inc., Dublin, CA), conducted prior to any experimental tests, to ensure adequate perimetric experience. False positive rates were < 10% for all participants. All experimental tests were conducted with natural pupils. A half-eye trial frame was worn, and full aperture trial lenses were used to correct refractive error for a viewing distance of 30 cm and/or induce optical defocus as required. The non-test eye was occluded with a patch. Details of participants recruited to each experiment are given in sections *6.3.1* and *6.4.1*

6.3   Experiment One – The effect of optical defocus[1]

As discussed in section *6.1.1*, the effects of optical defocus are largely unavoidable in perimetric testing. As such, it is crucial that these effects are fully quantified with any new stimulus form, and compared with the current reference standard (Goldmann III). Here, an experiment was conducted to determine the effects of optical defocus on the three stimulus forms optimised to probe changing spatial summation in glaucoma (A, AC, and C_R stimuli) in comparison with a Goldmann III equivalent (GIII) stimulus, although for the purposes of this experiment, only healthy participants were recruited. Atchison (1987) identified four factors that influence the relationship between optical defocus and visual field measurement: stimulus area, test eccentricity, background luminance, and defocus correction. Test eccentricity, background luminance, and defocus correction were all controlled for, in an attempt to isolate the effects of optical defocus on the four stimulus forms, permitting accurate conclusions to be drawn.

6.3.1   Methods

6.3.1.1 Apparatus and set-up

Parameters for each of the four stimulus forms modulated as described in chapter four (section *4.2.3*). As in previous experiments, all stimulus steps were converted to

---

[1]Data for this experiment was collected with the aid of Katherine Ward, as part of her undergraduate research project.

energy according to *Equation 3.3*, and equated in terms of log energy, to allow for a direct comparison between the different stimulus forms.

The same four visual field locations were tested, 9.9° from fixation along the 45°, 135°, 225°, and 315° meridians, as in chapter four (*Figure 6.1*). In using locations equidistant from fixation, any differences in the effects of optical defocus due to eccentricity are therefore controlled for. This does not control for peripheral aberrations induced by spectacle lenses, either due to the participant's own refractive correction, or due to the additional positive power used to induce blur, however the peripheral aberrations, and magnification/minification effects, may be expected to be constant between all stimulus forms. Although a differing refractive error is expected at peripheral compared to foveal locations, there is general agreement with respect to the effects of optical defocus between studies that have not corrected peripheral refractive error (Sloan 1961; Fankhauser and Enoch 1062; Benedetto and Cyrlin 1985; Campbell and Green 1965; Atchison 1987), and those that have (Anderson et al. 2001). It has also been shown that peripheral refractive error does not differ substantially from that at the fovea within the central ten degrees, i.e. the test locations used in this experiment (Ferree et al. 1931; Ferree and Rand 1933; Rempt et al. 1971; Millodot 1981). The decision was made to correct the foveal refractive error only, as this would be typical of the procedure undertaken in a clinical setting, and largely due to the inability to correct all peripheral and central refractive errors simultaneously.
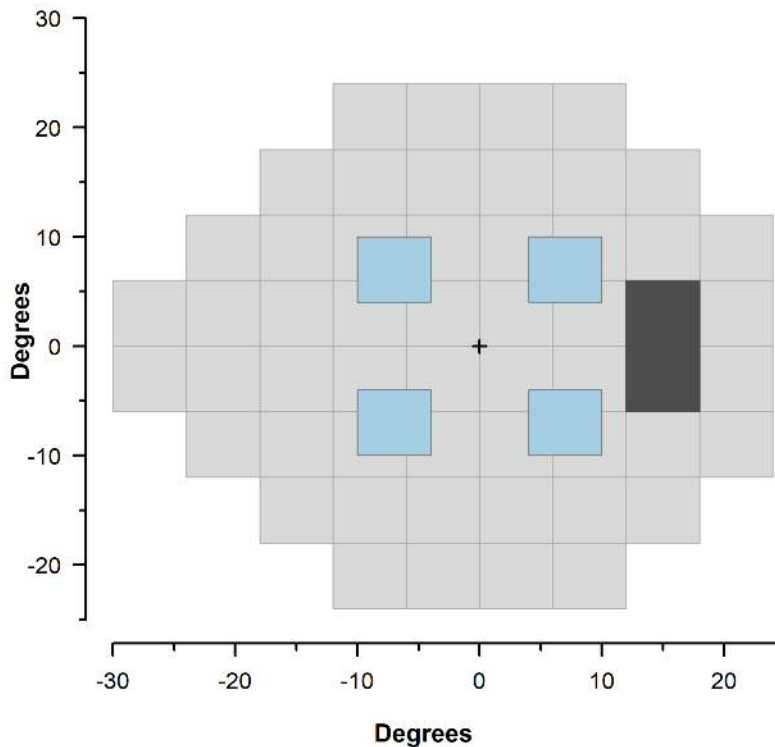
Figure 6.1 – The four test locations used in this experiment (shown in blue), in relation to a standard 24-2 test grid (shown in grey), for a right eye.

6.3.1.2 Psychophysical procedure

The same three-stage protocol as used in chapter four, utilising a MOCS procedure was adopted here. This is described fully in section *4.2.4*, but is briefly recapped as follows:

1. An adaptive, 1:1 staircase procedure of six reversals determined an approximate threshold at each of the four test locations (randomly interleaved). This was repeated twice, and the threshold was taken as the mean of the final four reversals from the second test. A rest break was taken between the two tests.

2. A short MOCS procedure was undertaken at each of the four test locations, using nine energy levels, presented five times each (180 presentations in total). These energy levels were the threshold established from the staircase procedure, and six further values, three above and three below this threshold (±0.45 log energy in 0.15 log energy intervals). Two additional energy levels were established as the threshold from the staircase procedure ±0.9 log

energy. Test locations were randomly interleaved, and stimulus presentations were randomised by energy level. A rest break was taken halfway (after 90 presentations). A FOS curve was constructed from the participant responses, and was fitted with a psychometric function. Energy values at p(seen) = 0.1, 0.3, 0.5, 0.7, and 0.9 were estimated from the psychometric function.

3. A standard MOCS procedure was undertaken at each of the four test locations, using eight energy levels, presented 20 times each (640 presentations in total). Five of the energy levels were p(seen) = 0.1, 0.3, 0.5, 0.7, and 0.9 from the short MOCS procedure, with three additional energy levels at p(seen, 0.5) ± 2*SD of the psychometric function (according to *Equation 3.2*), and p(seen, 0.5) + 1.5 log energy. Test locations were randomly interleaved, and stimulus presentations were randomised by energy level. A rest break was taken at every quarter (after 160 presentations). A FOS curve was constructed from participant responses and fitted with a logistic psychometric function. Threshold was established as p(seen) = 0.5, and response variability was established as the SD of the psychometric function (according to *Equation 3.2*).

Participants could request additional rest breaks as required.

6.3.1.3 Participants

Three young, healthy participants aged 20.7, 20.8 and 21.3 years were recruited to this experiment. The right eye was selected as the test eye for all three participants.

One participant was emmetropic, and two participants were myopic (-3.50/-0.25 x 180, and -2.25/-0.25 x 160). MD values with the SITA Standard 24-2 program on the HFA II were -1.09 dB, -1.15 dB, and -1.18 dB.

6.3.1.4 Experimental phases

Two phases were conducted as follows:

Phase one

To determine the effect of blur on inter-test variability, one emmetropic participant completed tests for the GIII stimulus six times – three times with a +3.25 DS working distance correction (i.e. baseline), and three times with an additional +4.00 DS; these

six tests were conducted in random order. Threshold, i.e. p(seen) = 0.5, and response variability (according to *Equation 3.2*) from the third test stage (i.e. the standard MOCS procedure), were compared within each category of optical defocus. In addition, this stage was used to confirm that +4.00 DS was sufficient to induce a statistically significant threshold increase with a Goldmann III equivalent stimulus.

Phase two

The other two myopic participants completed tests for all four stimulus forms under three conditions of optical defocus. One set of tests was undertaken with participants' full, distance refractive correction in place, with the addition of a +3.25 DS working distance correction (i.e., baseline), another set of tests was undertaken with an additional +2.00 DS, and a further set of tests was undertaken with an additional +4.00 DS. Stimulus forms, and levels of optical defocus, were randomised for each participant. Participants completed all tests within 15 days.

Ethical approval for the experiment was obtained from the School of Optometry and Vision Sciences Research and Audit Ethics Committee, Cardiff University. The research adhered to the tenets of the Declaration of Helsinki. Written, informed consent was obtained from all participants prior to inclusion.

6.3.1.5 Statistical Analysis

Fitting of the psychometric functions, and analysis of FOS data was performed in MATLAB (version R2015b; The MathWorks Inc., Natick, MA, USA), using the Palamedes toolbox (Prins and Kingdom 2009). Analyses described from this point were conducted on those FOS data collected in stage three (the standard MOCS procedure). Further statistical analyses were performed with the open source statistical environment R (R Development Core Team, 2017).

Phase one

To evaluate the repeatability of the results obtained, threshold (p(seen) = 0.5), and response variability (SD) were pooled across the four test locations, and analysed across the three tests, under each condition of optical blur for the single participant. A linear mixed effects analysis was performed of the relationship between threshold and test number, and response variability and test number, using the lme4 package (Bates

et al. 2015). Test number and condition of optical focus (i.e. baseline, or with +4.00 DS optical blur), without an interaction term, were entered as fixed effects. Intercepts for test locations were entered as random effects. Data were analysed to ensure they met the necessary assumptions, namely a normal distribution and a lack of heteroskedasticity; where results of a linear mixed effects analysis have been reported, there were no violations of these assumptions.

P-values were obtained by likelihood ratio tests of the full model with the effect in question (i.e. test number), against the model without the effect in question. This was repeated for response variability. Two-tailed p-values are quoted for analyses of repeatability.

To evaluate the effect of increased optical defocus on threshold and response variability, a linear mixed effects analysis of the relationship between threshold and optical focus, and response variability and optical focus, was conducted on pooled data from all four test locations. As the introduction of additional optical blur will result in an increase in threshold only, one-tailed p-values have been quoted here for the analyses of threshold with increasing optical defocus. This is not true for response variability, which may increase or decrease with the addition of optical blur; two-tailed p-values are therefore quoted for analyses of response variability with increase in optical defocus.

Phase two

Threshold (p(seen) = 0.5) and response variability (SD) were pooled across the four locations for the two participants, and analysed across the three conditions of optical focus. A linear mixed effects analysis of the relationship between response variability and stimulus form was performed, and p-values are quoted from a likelihood ratio test.

To investigate the effect of optical defocus on stimulus form, threshold increase from baseline for +2.00 DS, and +4.00 DS, was determined and analysed with a linear mixed effects model. Two-tailed p-values are quoted for comparisons between stimuli.

This analysis was also conducted individually for each stimulus form, to examine the effect of optical defocus on threshold and response variability within a single stimulus. One-tailed p-values are quoted for analyses of threshold difference, and two-tailed p-values are quoted for analyses of response variability, with increasing optical defocus.

SNR was also investigated for each of the four stimulus forms. TD was established as the difference between measured threshold and expected threshold. Expected threshold for the participant's age was established from the linear regression models from section *4.3*, which modelled the relationship between threshold energy and age in the healthy participants recruited to the experiment presented in chapter four (*Figure 6.2*).
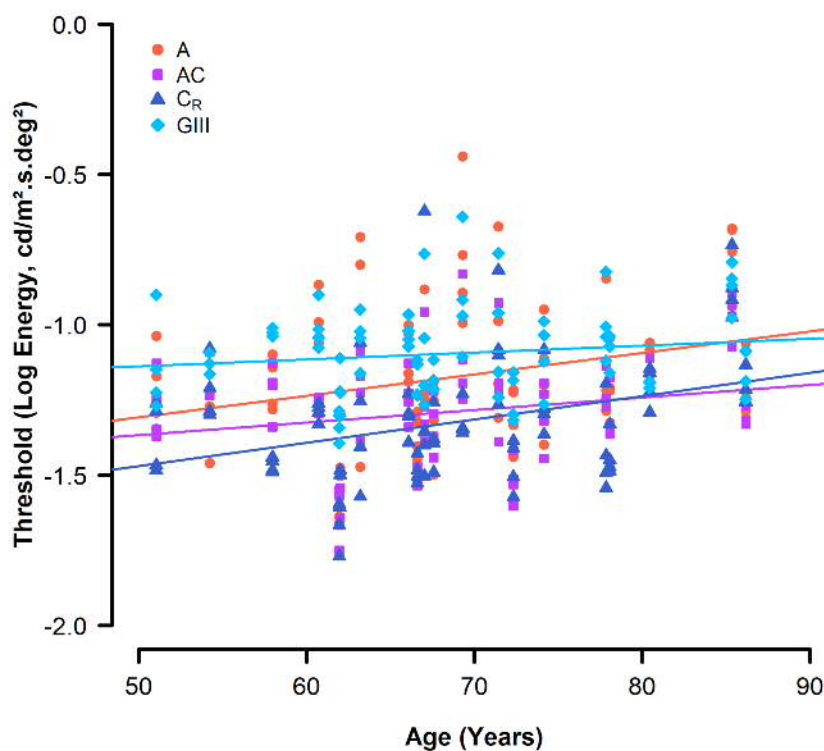
Figure 6.2 – Threshold energy, pooled for the four test locations, plotted against age and fitted with a mixed model linear regression.

SNR was then established as TD/response variability, as in chapter four. The SNR difference for +2.00 DS and +4.00 DS from baseline was determined, and analysed with a linear mixed effects model as described above, to evaluate SNR difference from

baseline with stimulus form. Two-tailed p-values are quoted for comparisons between stimuli.

This analysis was also conducted individually for each stimulus, to examine the effect of optical defocus on SNR within a single stimulus form. One-tailed p-values are quoted for analyses of SNR with increase in optical defocus.

In all statistical analyses, a Holm-Bonferroni post hoc correction was applied where multiple tests of the same hypothesis occurred. All p-values quoted have been post hoc corrected.

6.3.2   Results

6.3.2.1 Phase one

Figure 6.3 shows *(A)* threshold and *(B)* response variability for the GIII stimulus at baseline, and with an additional +4.00 DS, repeated three times each for one participant. Results for the four test locations are shown, with horizontal jitter added for ease of data visualisation. This figure indicates an increase in threshold with the addition of +4.00 DS, but no apparent difference in response variability. This was confirmed by the linear mixed effects analysis, which indicated a statistically significant increase in threshold with the additional +4.00 DS compared with baseline, by 0.74 cd/m^2.s.deg^2 ± 0.03 SE (p < 0.01), but no statistically significant difference in response variability (p = 0.10).
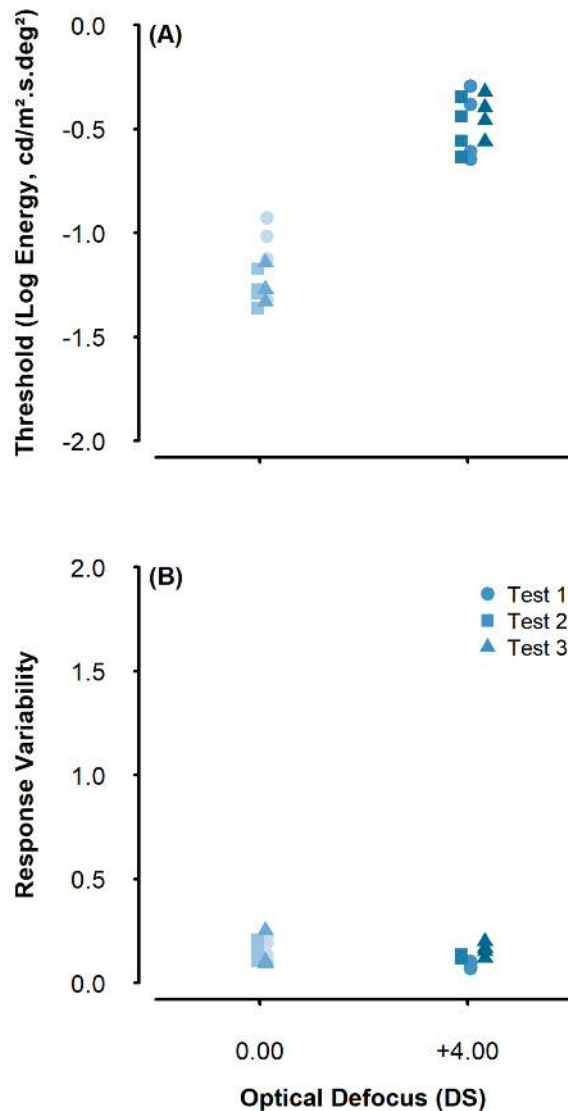
Figure 6.3 – (A) Threshold and (B) Response variability for a GIII stimulus for the four test locations, at baseline, and +4.00 DS optical blur, repeated three times each. Horizontal jitter has been added for ease of data visualisation.

Repeatability of the three tests was also investigated, with statistical tests conducted individually for each of the two conditions of optical defocus. No statistically significant differences were found between the three tests, either for baseline, or for +4.00 DS; this was true for both threshold, and response variability ($p > 0.05$).

This phase established the repeatability of the test procedure, and confirmed that an addition of +4.00 DS was sufficient to induce a statistically significant increase in threshold with a Goldmann III equivalent stimulus.

6.3.2.2 Phase two

<u>Threshold</u>

Figure 6.4 shows the threshold values obtained for each of the four stimulus forms, pooled across all four test locations for the two participants. Data are shown for each condition of optical defocus. *Figure 6.4.A* shows the raw threshold values, and *Figure 6.4.B* shows the threshold difference from baseline. From this figure, the A and GIII stimuli appear to have the highest raw threshold values for each of the three conditions of optical focus (*Figure 6.4.A*), but also appear to show the lowest increase in threshold with the addition of optical blur (*Figure 6.4.B*).

A comparison between the four stimulus forms, analysing threshold increase from baseline for both +2.00 DS and +4.00 DS, did not indicate a statistically significant difference between stimulus forms ($p = 0.06$).

Further analysis, conducted individually for each stimulus, confirmed that the addition of +2.00 DS resulted in a statistically significant threshold increase for all four stimulus forms (all $p < 0.01$), and the addition of a further +2.00 DS (i.e. the difference between +2.00 DS and +4.00 DS) resulted in a further statistically significant threshold increase for all four stimulus forms (all $p < 0.01$).
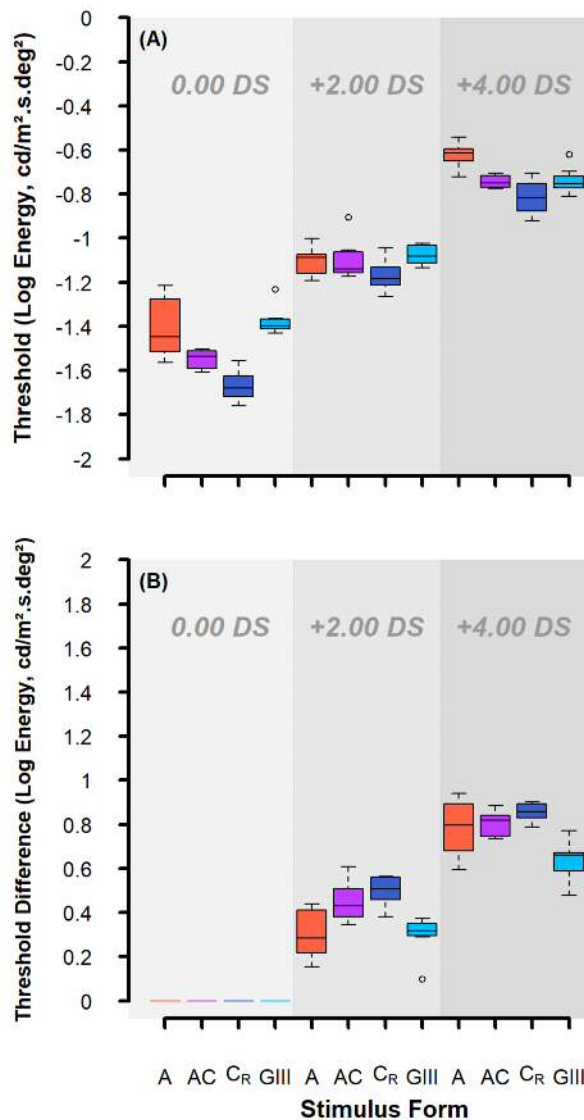
Figure 6.4 – Threshold values for each of the four stimulus forms, for each of the three conditions of optical blur, pooled across the four tested locations. (A) Raw threshold values, and (B) Threshold increase from baseline.

Response variability

Figure 6.5 shows the response variability values obtained for each of the four stimulus forms, pooled across all four test locations for the two participants. Data are shown for each condition of optical defocus. *Figure 6.5.A* shows the raw values for response variability, and *Figure 6.5.B* shows the difference in response variability from baseline. There is no clear association between response variability and optical defocus; a greater response variability is observed with an increase in optical defocus with some stimuli, and a lower response variability is observed with an increase in optical defocus

with other stimuli. This is not consistent across the two levels of optical defocus, e.g. with the A, AC and C_R stimuli, a lower response variability is observed with +2.00 DS, but a slightly higher response variability is observed from +2.00 DS to +4.00 DS.
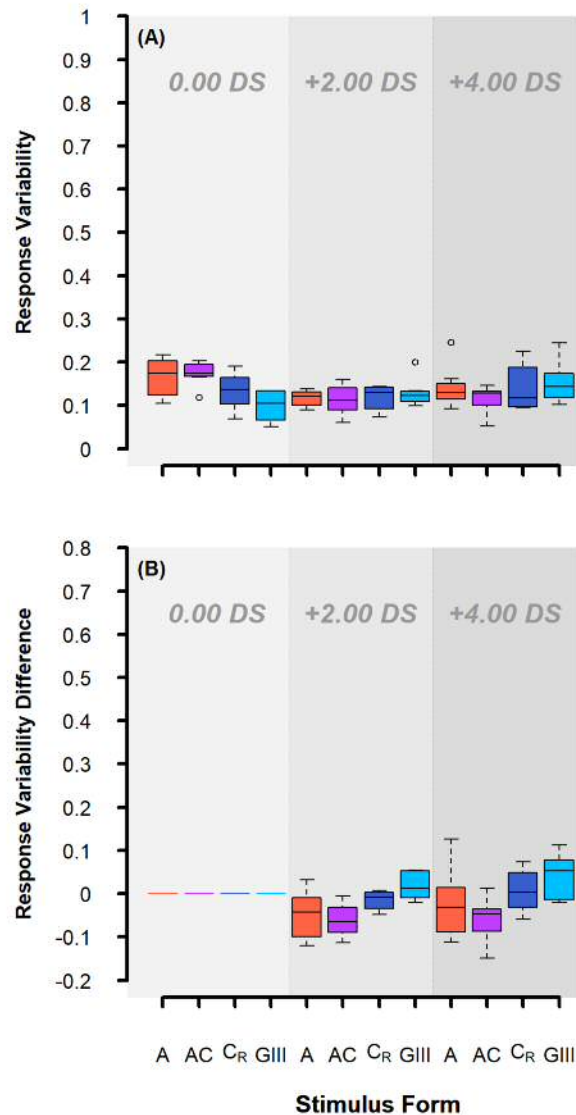
Figure 6.5 – Response variability for each of the four stimulus forms, for each of the three conditions of optical defocus, pooled across the four test locations. (A) Raw response variability, and (B) Response variability difference from baseline.

The level of optical defocus had a statistically significant effect on response variability (p = 0.01), when considering the complete linear mixed model. Further analysis, conducted individually for each stimulus form, identified a statistically significant

difference in response variability due to the level of optical defocus with the AC stimulus only (p = 0.001); differences in response variability were not statistically significant with level of optical defocus for the other three stimulus forms (all p > 0.05). A pairwise comparison of the levels of optical defocus for the AC stimulus, found the response variability to be statistically significantly higher at baseline compared with +2.00 DS, and +4.00 DS (both p < 0.01); no statistically significant difference was found between response variabilities for +2.00 DS and +4.00 DS.

A comparison between the four stimulus forms, including response variability for all three levels of optical focus, did not indicate any statistically significant differences between stimulus forms (p = 0.95).

<u>Signal/noise ratio</u>

Figure 6.6 shows the SNR for the four stimulus forms, pooled across all four test locations, and shown for each condition of optical defocus. A positive SNR at baseline indicates a higher threshold measurement than estimated for participants' age from the linear regression model, whilst a negative SNR at baseline indicates a lower threshold measurement than estimated. *Figure 6.6.A* shows the raw SNR values, and *Figure 6.6.B* shows the SNR difference from baseline.

Raw SNR values appear to be lower with the GIII stimulus in comparison with the other three stimulus forms for all conditions of optical defocus (*Figure 6.6.A*). SNR appears to show a lower increase with increasing optical defocus for the GIII stimulus, compared with the other three stimulus forms, particularly with +4.00 DS. However, a comparison of SNR difference from baseline with both +2.00 DS and +4.00 DS did not indicate a statistically significant difference between stimulus forms (p = 0.13).
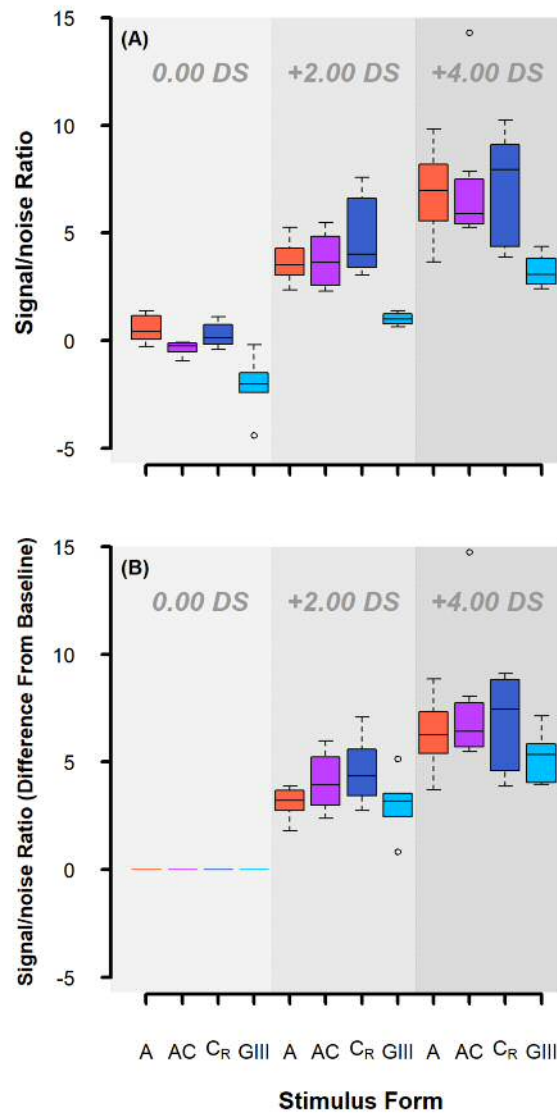
Figure 6.6 – SNR for each stimulus form, for each of the three conditions of optical defocus, pooled across the four test locations. (A) Raw SNR, and (B) SNR increase from baseline.

Further analysis, conducted individually for each stimulus form, confirmed that the addition of +2.00 DS resulted in a statistically significant SNR increase for all four stimulus forms (all p < 0.01), and the addition of a further +2.00 DS (i.e. the difference between +2.00 DS and +4.00 DS) resulted in a further statistically significant SNR increase for all four stimulus forms (all p < 0.01).

6.4   Experiment Two – The effect of intraocular straylight

As discussed in section *6.1.2*, the effects of intraocular straylight are an important consideration in perimetry. Given the age demographic of patients typically undertaking such tests, lens opacities such as cataract are commonplace, in addition to subtler, age-related lenticular changes. As such, it is crucial that these effects are fully quantified with any new stimulus form, and compared with the current reference standard (Goldmann III). Here, an experiment was conducted to determine the effects of straylight on the three stimulus forms optimised to probe changing spatial summation in glaucoma (A, AC, and C_R stimuli), in comparison with a Goldmann III equivalent (GIII) stimulus, although for the purposes of this experiment, only healthy participants were recruited.

6.4.1   Methods

6.4.1.1 Apparatus and set-up

Five white, opacity containing filters (Fog 1-Fog 5 Standard; LEE Filters, Andover, UK) were used to induce additional straylight. These filters have been used in previous studies to simulate differing amounts of lens opacification, to investigate the effects of straylight on various perimetric techniques (Zlatkova et al. 2006; Anderson et al. 2009; Bergin et al. 2011). They have been shown to simulate wide-angle scatter, as found in cataract, with the amount of straylight varying inversely proportionally to the square of the angular distance, a typical relationship found in intraocular light scatter (Ijspeert et al. 1990; Zlatkova et al. 2006).

Straylight was measured for each participant, under each of the six conditions (i.e. with refractive correction only, hereafter referred to as 'baseline', and with the addition of each of the five fog filters), quantified with the use of the C-Quant straylight meter (Oculus, Wetzlar, Germany). Visual acuity was not measured with the fog filters, as this has proved a poor predictor of straylight values (De Waard et al. 1992).

The C-Quant is a commercially available instrument, utilising the 'compensation comparison method' (Franssen et al. 2006), developed from the well-established 'direct compensation method' proposed by Van den Berg (1986). With this instrument, the observer monocularly fixates a central circular field, divided into two

halves (*Figure 6.7*). An outer, flickering ring forms the straylight source which flickers in phase with one of the central hemifields (test field B).
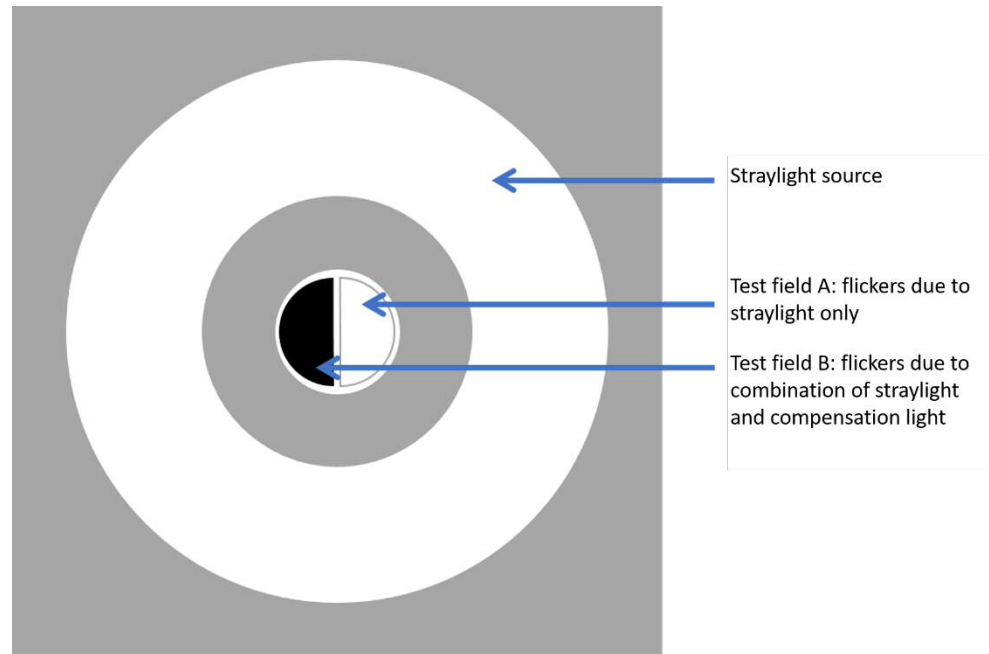
Figure 6.7 – Illustration of observer view as used in the C-Quant straylight meter (adapted from Franssen et al. 2006). Test fields are randomised throughout the test.

Due to this straylight, the observer perceives a superimposed flicker in the other half of the central field (test field A); there is no actual intrinsic flicker in this half of the field. The C-Quant then modulates the light in test field B in counterphase to the straylight source, which 'compensates' for the straylight flicker; in this field the observer therefore perceives a combination of the straylight (as also observed in the other half of the field), and the compensating, counterphase light. The task is a 2AFC, in which the observer is required to indicate which field appears to be flickering more strongly, by pressing the button corresponding to that field. The C-Quant then adjusts the amount of compensating, counterphase light for the subsequent presentation, using a maximum likelihood estimation (Van den Berg et al. 2005). The observer's responses are used to determine the threshold, i.e. the point at which both sides of the field appear to be flickering equally strongly, and thus observer responses are due to random chance only. This threshold is an indication of how much compensating,

counterphase light was necessary to neutralise the flicker perception from the straylight.

The test consists of approximately 25 presentations, with a test duration of 1.5-2 minutes, and is reportedly easy to explain and intuitive to observers, resulting in a highly repeatable test with ≤ 1% false positives/negatives (Franssen et al. 2006).

An output plot showing the estimated psychometric function is provided, along with the resulting threshold log straylight value. Threshold is also displayed relative to the instrument's age-related normative database. Two reliability indices are also included in the output plot, the estimated SD of the straylight threshold (ESD), and the quality factor for psychometric sampling (reliability coefficient, Q). In this experiment, in accordance with manufacturer's guidelines, if ESD was greater than 0.08, or Q was greater than 1, the test was discarded, and repeated.

Different ranges are available to use; if a higher amount of straylight is present, the range can be increased to enable a more accurate measurement. In this experiment, the standard range, 'E', was used for baseline measurements, and the three least dense filters (Fog 1-3), and a higher range, 'G', was used for the two densest filters (Fog 4-5), which increase straylight the most.

6.4.1.2 Participants

To determine the repeatability of the C-Quant measurements under each of the straylight conditions, six young, healthy participants underwent repeated testing with this instrument. Median [IQR] age was 26.7 years [26.1, 28.7]. Three participants had brown irides, two had blue irides, and one had hazel irides. Three right eyes and three left eyes were tested. Five participants were myopic (median [IQR] sphere: -2.55 DS [-3.00, -1.50]), and one was hyperopic (+4.00 DS), with median [IQR] cylinder of -0.75 DC [-0.94, -0.38].

To determine the effect of straylight on each of the four stimulus forms, five young, healthy participants, all of whom had previous psychophysical experience, were recruited. Median [IQR] age was 29.6 years [27.1, 31.4]. Median [IQR] MD was -0.24

dB [-0.51, 0.28] with the SITA Standard 24-2 program on the HFA II. Four right eyes and one left eye were tested.

One participant was emmetropic and the other four participants were myopic (median [IQR] sphere: -2.75 DS [-3.44, -2.19], median [IQR] cylinder: -0.75 DC [-1.06, -0.56]). As all participants were pre-presbyopic, with no known accommodative issues, no near add was incorporated.

6.4.1.3 Test procedure

<u>Repeatability of C-Quant straylight measurements</u>

To investigate the repeatability of the C-Quant straylight meter, six participants completed the straylight measurement tests three times for each of the six straylight conditions. The three measurements were taken consecutively, and in accordance with the manufacturer's guidelines of eye-positioning in relation to the instrument. The order of the straylight conditions were randomised for each participant, and participants wore their full refractive correction. The five fog filters were cut as trial lenses, and fitted into the trial frame worn. Care was taken to ensure that all lenses were clean, to avoid anomalous results. All measurements were taken on the same day, with rest breaks between each test.

Three of these six participants, plus two further participants who had not performed the repeatability tests, completed the experiment evaluating the effects of straylight on threshold for each of the four stimulus forms, as described below.

<u>Evaluating the effects of straylight on threshold</u>

As the five fog filters were flat, and were not subject to the effects of peripheral defocus as found with corrective lenses, locations at various eccentricities could be more readily utilised than in the experiment on optical defocus; participants' own refractive correction may influence peripheral defocus, but as this remains constant in all tests it should not be a confounding factor. As such, the 18 visual field locations as used in chapter five were tested here. An adaptive, 1:1 staircase of four reversals was used at each of the 18 test locations (randomly interleaved), as described in more detail in chapter five (section *5.2.2.3*). Participants completed tests for each of the four stimulus forms under each of the six straylight conditions, once at baseline (i.e.

with refractive correction only), and once with the addition of each of the five fog filters. As in chapter five, tests in which the false positive rate was higher than 25% were discarded, and repeated. The order of the tests was randomised for each participant. Participants were given a rest break between tests, and could request additional rest breaks at any time. Tests were completed within a two-month period.

Ethical approval for the experiment was given by the East of Scotland Research Ethics Committee (NHS Scotland). The research adhered to the tenets of the Declaration of Helsinki. Written, informed consent was obtained from all participants prior to inclusion.

6.4.1.4 Statistical analysis

Statistical analyses were performed with the open source statistical environment R (R Development Core Team, 2017), and SPSS (IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0, Armonk, NY: IBM Corp). To ensure appropriate comparisons between quadrants, all data were converted to that for a right eye.

6.4.1.5 Repeatability of C-Quant straylight values

To examine the repeatability of the C-Quant straylight measurements, a one-way ANOVA with repeated measures was conducted to compare the effect of test number on log straylight; this was conducted separately for each straylight condition. Mauchly's Test of Sphericity was used to determine whether assumptions of sphericity were violated; where this assumption was violated, a Greenhouse-Geisser correction was applied.

As there is some indication that ocular pigmentation influences straylight, with lighter coloured eyes reportedly having higher levels of straylight than darker eyes (Van den Berg et al. 1991; Van den Berg 1995), an analysis was conducted to determine whether ocular pigmentation had a statistically significant effect on baseline straylight values. A Mann-Whitney U test was conducted for this analysis. There is also some indication that refractive error status influences straylight, with myopic eyes reportedly having higher levels of straylight than hyperopic eyes (Rozema et al. 2010), however this could not be tested statistically in this experiment, given that five out of the six participants were myopic, and only one participant was hyperopic.

To evaluate inter-observer variability, the mean straylight value for each straylight condition was determined for each participant, and a one-way ANOVA conducted. Levene's Test was used to determine whether assumptions of variance homogeneity were violated; where this assumption was violated, a Welch ANOVA was performed as an alternative.

6.4.1.6 Effect of straylight

To evaluate the effect of straylight on each of the four stimulus forms, threshold values from the 18 test locations were pooled for each stimulus, under each straylight condition. A linear mixed effects analysis was performed of the relationship between threshold difference from baseline with the five fog filters, and stimulus form, using the lme4 package (Bates et al. 2015). Stimulus form and straylight condition, without an interaction term, were entered as fixed effects. Intercepts for subjects and test locations, as well as by-subject random slopes for the effect of stimulus form, were entered as random effects. P-values were obtained by likelihood ratio tests of the full model with the effect in question (i.e. stimulus form), against the model without the effect in question. Two-tailed p-values are quoted for comparisons between stimulus forms.

A further analysis evaluated the relationship between threshold and straylight condition, conducted individually for each stimulus form. Pairwise comparisons were carried out between baseline threshold, and threshold with each of the five fog filters, to determine when threshold increase from baseline with increase in straylight first became statistically significant. As the introduction of additional straylight will result in an increase in threshold only, one-tailed p-values have been quoted here.

In addition to considering all test locations together, the 18 test locations were subdivided into five zones of eccentricity, to determine whether the impact of straylight on threshold differed with distance from fixation. The test locations included in each of these five zones are shown in *Figure 6.8*. The linear mixed effects analyses, as described above, were also carried out on test locations within each of these five zones.

In all statistical analyses, a Holm-Bonferroni post hoc correction was applied where multiple tests of the same hypothesis were undertaken. All p-values quoted here have been post hoc corrected.
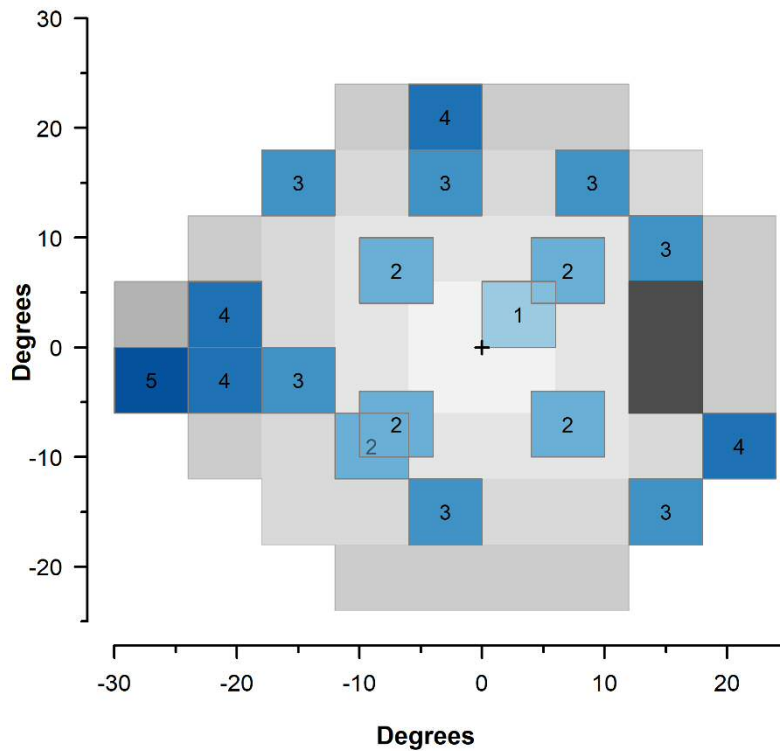
Figure 6.8 – Subdivision of the 18 test locations into five zones according to their eccentricity from fixation. Locations with the same number were included in the same zone.

6.4.2   Results

6.4.2.1 Repeatability of C-Quant straylight values

All participants completed the straylight measurements accurately under all conditions of straylight; ESD and Q measurements did not indicate a need to repeat any measurements.

Figure 6.9 shows the three log straylight values from the C-Quant straylight meter for each of the six participants, at baseline. These values are plotted against age, with the solid line indicating mean straylight value from the C-Quant's normative database, and the light grey band indicating the normative range from this database. It can be seen

that log straylight values were above the mean straylight value from the C-Quant's normative database for all participants; some participants were within the normative range of the C-Quant's database, and some were not.
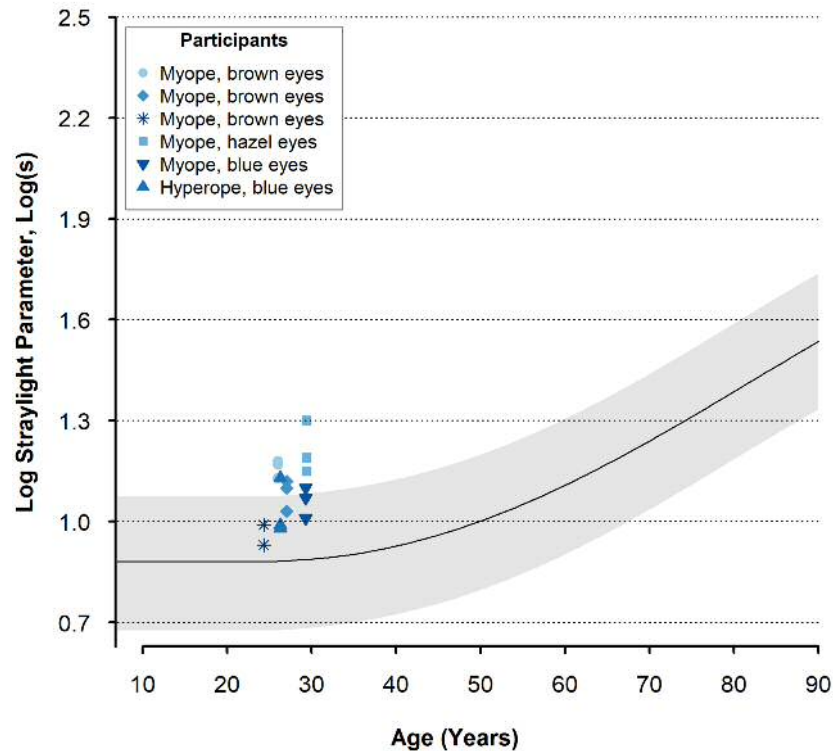
Figure 6.9 – Log straylight values for six participants at baseline. Measurements were repeated three times each using the C-Quant straylight meter, and plotted against age. The solid line denotes the mean straylight value, and the pale grey band denotes the normative range, from the C-Quant's normative database, as a function of age.

Each participant's refractive status and eye colour are also displayed in *Figure 6.9*. There does not appear to be a notable association between straylight values and either eye colour or refractive status. Mean log straylight values were slightly higher for participants with light coloured eyes (hazel eyes were included in this category), at 1.1 log(s), compared with 1.07 log(s) for participants with darker eyes, however the Mann-Whitney U test confirmed that there was no statistically significant difference in baseline straylight measurements between the two categories (p = 0.83). Mean log straylight values were slightly higher for the myopic participants (1.10) compared with the hyperopic participant (1.03).

Figure 6.10 shows the three log straylight values from the C-Quant straylight meter for each of the six participants, under each of the six straylight conditions. Mean straylight increase ranged from 24.5% with Fog 1 to 85.8% with Fog 5. Bergin et al. (2011), who tested six participants of a similar age to those recruited to this experiment, defined three categories of log straylight: (i) 0.6-1.2 log(s), classified as 'within normal limits', (ii) 1.2-1.6 log(s), classified as 'outside normal limits', representative of expected, normal aging changes, and (iii) 1.6-2.1 log(s), classified as 'significant cataract'. These categories are represented in *Figure 6.10* by the shaded areas; significant cataract is denoted by the darker red shaded area, outside normal limits is denoted by the lighter red shaded area, and within normal limits is denoted by the unshaded area below.

It is worth noting that baseline straylight values appear more variable, both intra- and inter- participant, in comparison with other straylight conditions. Indeed, one straylight value is outside normal limits at baseline. Fog filters 1 and 2 increase straylight such that all values are outside normal limits. For Fog 3, some values are outside normal limits, and some are consistent with significant cataract. Fog filters 4 and 5 increase straylight such that all log straylight values are consistent with significant cataract, in accordance with the categories defined by Bergin et al. (2011).
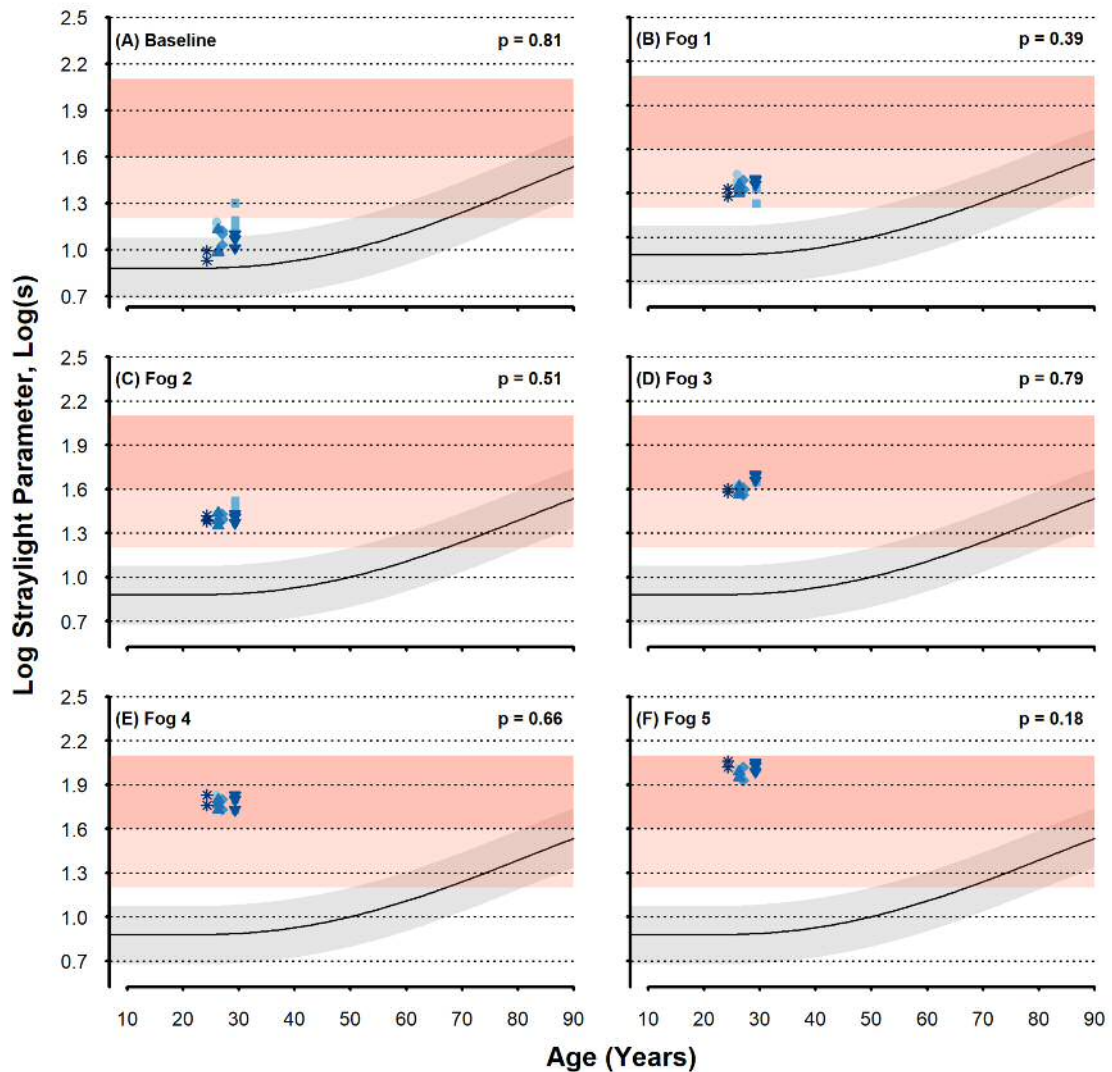
Figure 6.10 – Log straylight values for six participants, under six straylight conditions. Measurements were repeated three times each, using the C-Quant straylight meter, for each straylight condition (A-F), and plotted against age. The solid line denotes the mean straylight value, and the pale grey band denotes the normative range, from the C-Quant's normative database, as a function of age.

A one-way ANOVA with repeated measures found no statistically significant difference in log straylight value between the three tests; this was conducted separately for each of the straylight conditions, and p-values from this analysis are displayed in *Figure 6.10*. Mauchly's Test of Sphericity did not indicate a violation of the sphericity assumption within any straylight condition (all p > 0.05), so no Greenhouse-Geisser correction has been applied here.

A one-way ANOVA found no statistically significant difference between straylight values for the six participants (p = 0.999). Levene's Test did not indicate a violation of the assumption of variance homogeneity (p = 0.995), so a Welch ANOVA was not required.

6.4.2.2 Effect of straylight

Mean log straylight values for the five participants, who completed tests under the six straylight conditions, for each of the four stimulus forms, are shown in *Figure 6.11*. As in section *6.4.2.1*, all participants completed the straylight measurements accurately under all conditions of straylight; ESD and Q measurements did not indicate a need to repeat any measurements. The same point markers are used to denote individual participants in *Figure 6.11* as those used in *Figure 6.10*, but do not correspond to the same participant.
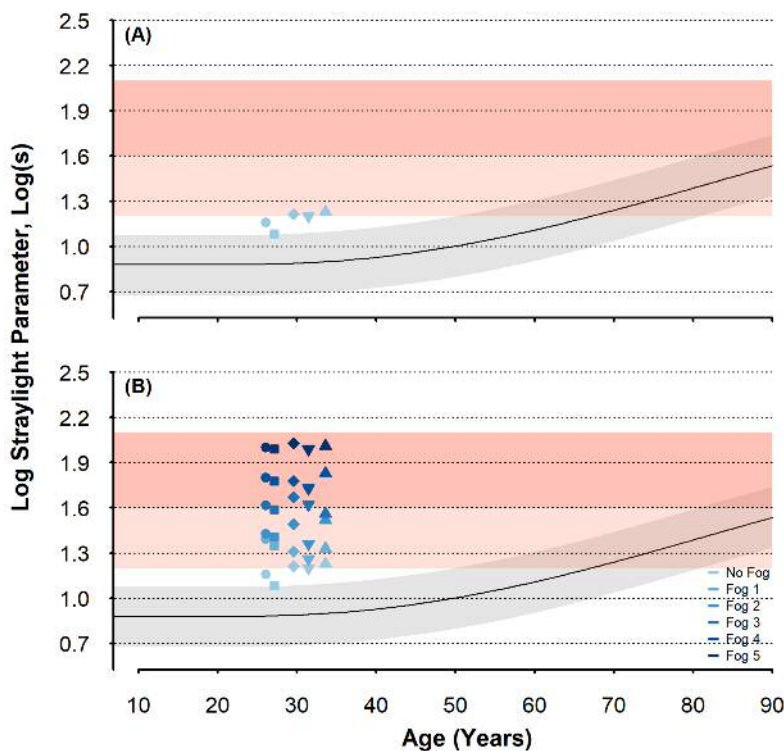
Figure 6.11 – Log straylight values, as measured using the C-Quant straylight meter, plotted as a function of age for each of the five participants (A) at baseline, and (B) under each of the six straylight conditions. The solid line denotes the mean straylight value, and the pale grey band denotes the normative range, from the C-Quant's normative database, as a function of age.

Baseline straylight values are shown in *Figure 6.11.A*, plotted against age; as in *Figure 6.10*, the solid line indicates the mean straylight values from the C-Quant's normative database, and the light grey band indicates the normal range from this database. The red shaded areas delineate the three straylight categories as described in section *6.4.2.1*. Log straylight values for all five participants were above average when compared with the C-Quant's normative database, and indeed some were classed as outside normal limits according to the previously outlined categories. However, as the straylight values were shown to be repeatable between measurements, and were not statistically significantly different between participants in section *6.4.2.1*, this was not considered an inaccuracy.

Figure 6.11.B shows the log straylight values for each of the five participants under all six straylight conditions. Log straylight is observed to increase proportionally with each fog filter, and is similar between participants. Similar to that observed in section *6.4.2.1*, straylight values for Fog filters 1 and 2 increase straylight such that all values are outside normal limits, for Fog 3, some values are outside normal limits, and some are consistent with significant cataract, and Fog filters 4 and 5 increase straylight such that all values are consistent with significant cataract.

Figure 6.12 shows threshold values for each of the four stimulus forms, averaged across the 18 test locations, and averaged across the five participants; these are plotted against mean log straylight for the five participants. *Figure 6.12.A* shows raw threshold and straylight values, and *Figure 6.12.B* shows the difference from baseline. Error bars indicate threshold SD, and horizontal jitter has been added for ease of data visualisation. The blue, dashed line indicates the maximum stimulus energy available for the C_R stimulus; maximum stimulus energy for all other stimulus forms were beyond the range of the axes displayed. Threshold does not appear to increase beyond the dynamic range of any of the four stimulus forms in this experiment; as such, any differences noted cannot be attributed to differing dynamic ranges between stimulus forms. It is difficult to definitively differentiate the four stimulus forms in either of the two plots, although the raw mean threshold appears higher with the A

stimulus with Fog 5 (-0.57) compared with the other stimulus forms (-0.81, mean of AC, C$_R$ and GIII stimuli) in *Figure 6.12*.

Analysis of the effect of stimulus form on threshold difference from baseline found no statistically significant difference between stimulus forms (p = 0.26).
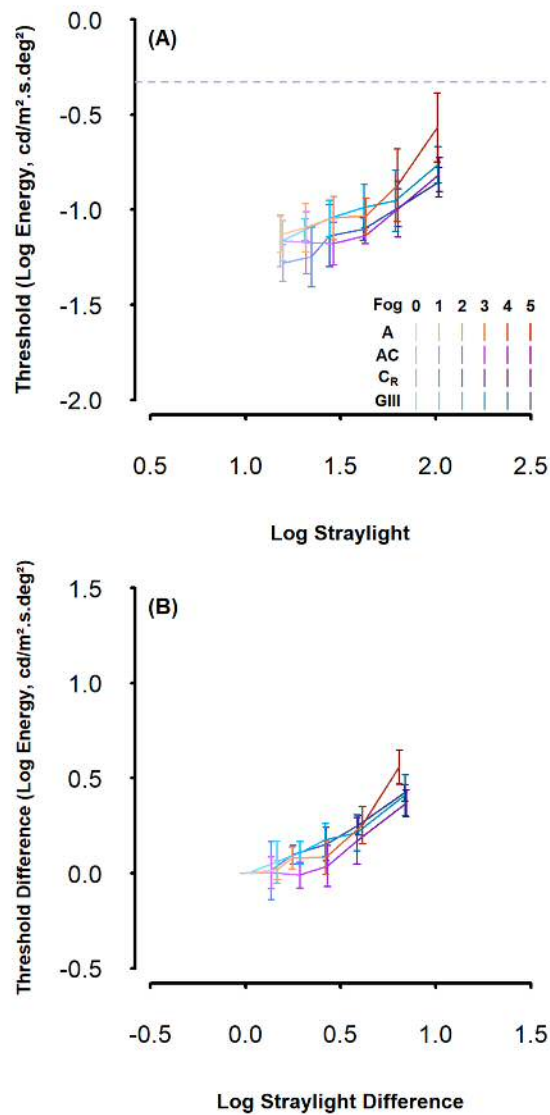
Figure 6.12 – Threshold values for each of the four stimulus forms, averaged across all 18 test locations, and averaged across the five participants. (A) Mean thresholds plotted against mean straylight values. (B) Mean threshold difference from baseline, plotted against mean straylight difference from baseline. Error bars indicate threshold SD, and horizontal jitter has been added for ease of data visualisation.

Analysis of the effect of straylight condition on threshold, conducted individually for each stimulus form, did indicate some differences between stimuli. For the GIII stimulus, a statistically significant increase in threshold was found with all five fog filters when compared with threshold at baseline ($p < 0.01$ for all comparisons). For the A and C_R stimuli, threshold increased from baseline with Fog 1, but this was not found to be statistically significant ($p = 0.44$, and $p = 0.28$ respectively); threshold increase from baseline with Fog 2-Fog 5 was statistically significant (all $p < 0.01$). For the AC stimulus, threshold increase from baseline was consistently the lowest increase in threshold across all straylight conditions; threshold increase with Fog 1-Fog 3 was not statistically significant (all $p > 0.05$), but threshold increase with Fog 4 and Fog 5 was found to be statistically significant (both $p < 0.01$).

Figure 6.13 shows threshold values within each of the five eccentricity zones (according to *Figure 6.8*), for each of the four stimulus forms. Threshold values have been averaged across the test locations in each zone, and across the five participants; these are plotted against mean log straylight. *Figure 6.14* shows the difference in threshold from baseline for each of the five fog filters, plotted against mean straylight difference from baseline. For both of these figures, as for *Figure 6.12*, error bars indicate threshold SD, and horizontal jitter has been added for ease of data visualisation. The blue, dashed line indicates the maximum stimulus energy available for the C_R stimulus. There appears to be a greater distinction between stimulus forms in zone 1 and, to a lesser extent, zone 2 (those closest to fixation) when observing raw values (*Figure 6.13*), although this does not hold true when observing differences from baseline (*Figure 6.14*).
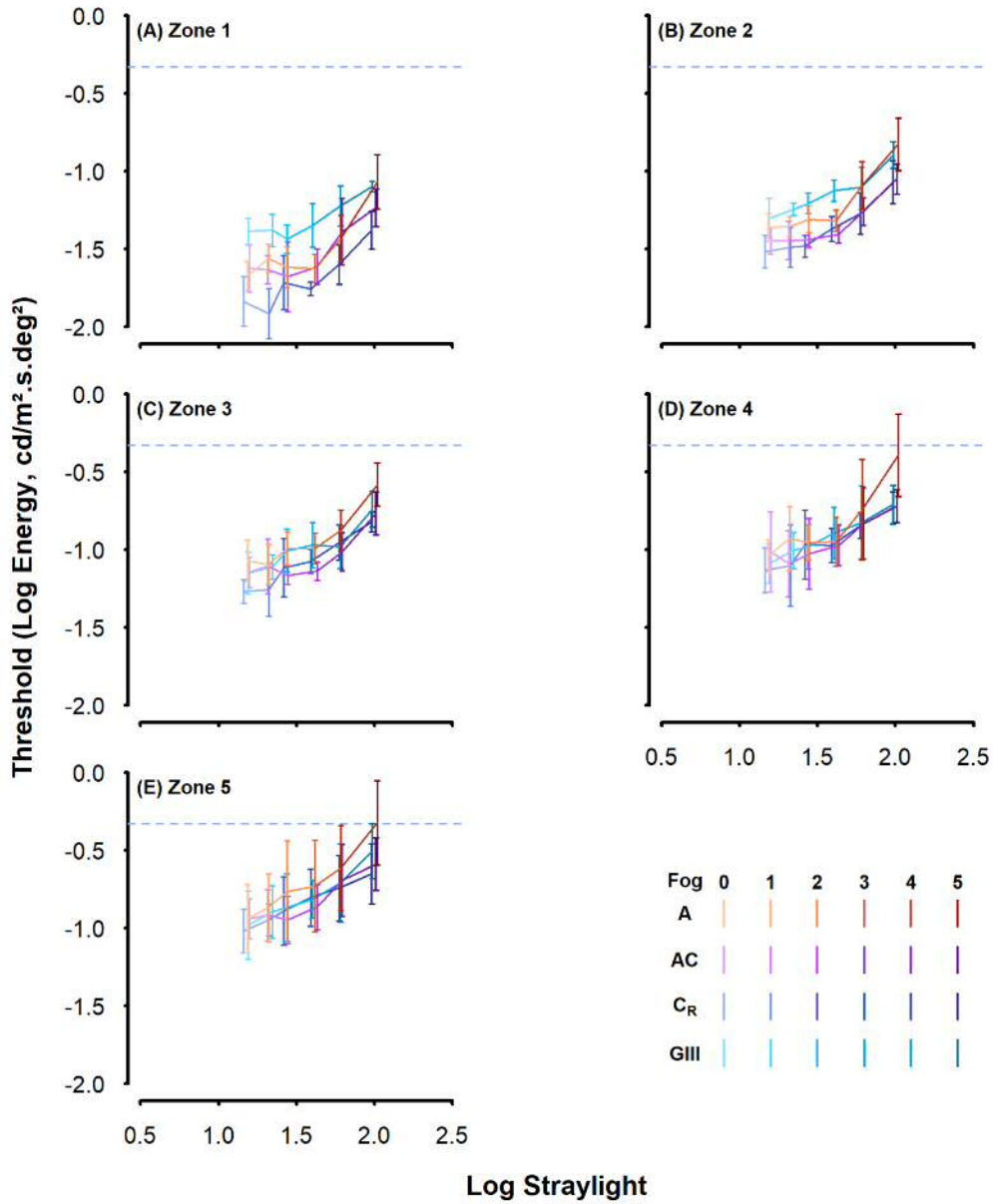
Figure 6.13 – Threshold values for each of the four stimulus forms, averaged within each of the five eccentricity zones, and plotted against mean straylight values. Error bars indicate threshold SD, and horizontal jitter has been added for ease of data visualisation.
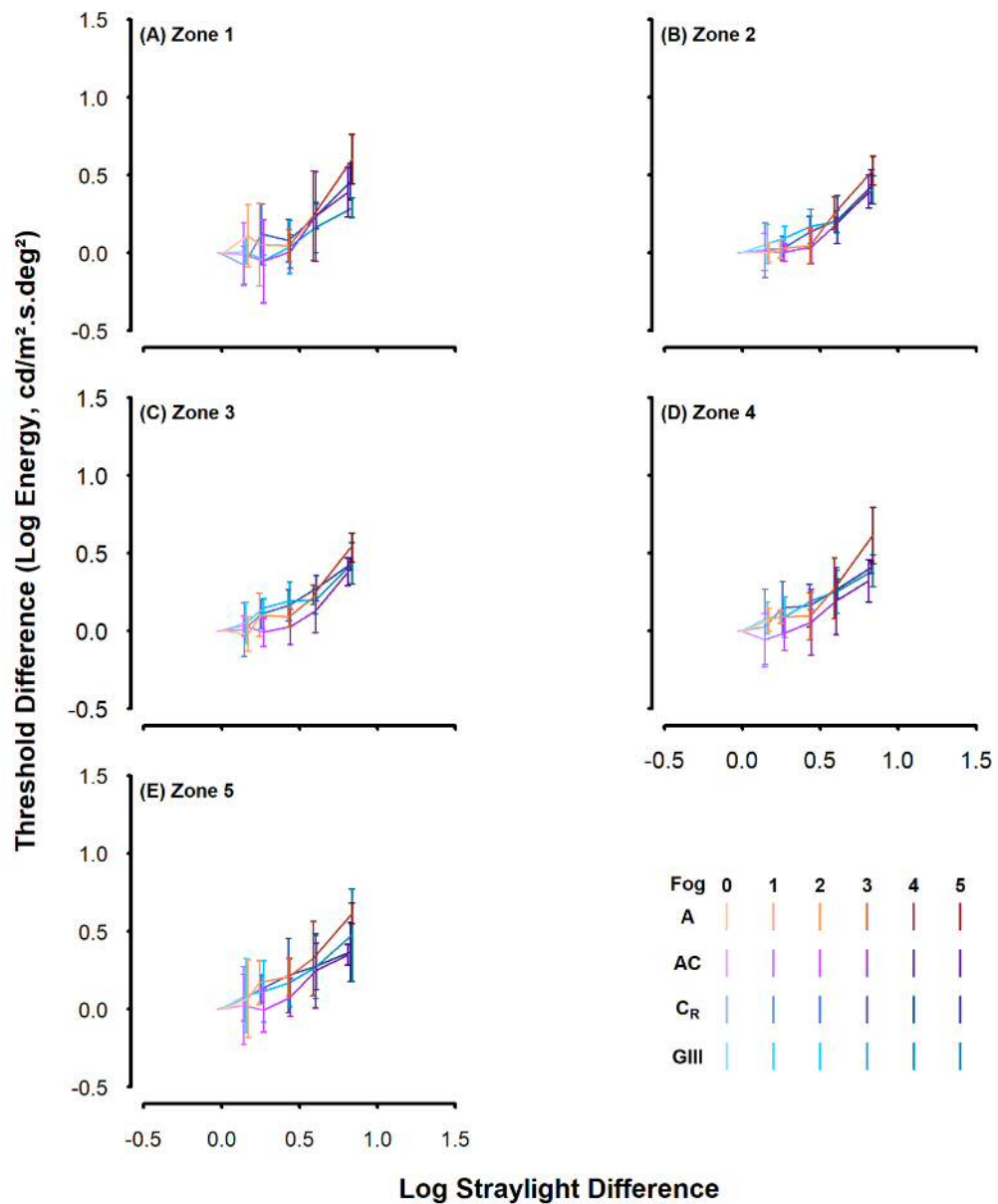
Figure 6.14 – Threshold difference from baseline for each of the four stimulus forms, averaged within each of the five eccentricity zones, and plotted against straylight difference from baseline. Error bars indicate threshold SD, and horizontal jitter has been added for ease of data visualisation.

Analysis of the effect of stimulus form on threshold difference from baseline found no statistically significant difference between stimulus forms, within any of the five eccentricity zones (all p > 0.05).

Analysis of the effect of straylight condition on threshold, conducted individually for each stimulus form within each of the five zones, did however indicate some differences between stimuli. *Table 6.1* indicates the fog filter at which threshold

increase from baseline first became statistically significant for each of the four stimulus forms, and within each of the five eccentricity zones.

Zone	A	AC	C_R	GIII
1	4	5	4	5
2	4	4	3	2
3	2*	4	2	2
4	4	4	3	3
5	4	5	4	5

Table 6.1 – The fog filter at which threshold difference from baseline first became statistically significant for each of the four stimulus forms, within each of the five eccentricity zones.

In zone 3, slightly anomalous findings are noted for the A stimulus, as indicated by the * in *Table 6.1*. In this zone, threshold increase from baseline for the A stimulus was statistically significant for Fog 2 (p = 0.04), but was not statistically significant for Fog 3 (p = 0.08). Threshold increase from baseline was statistically significant for Fog filters 4 and 5 (both p < 0.01). In all other analyses, threshold increase from baseline with all consecutive fog filters was found to be statistically significant (following the first fog filter to show a statistically significant threshold increase from baseline).

Observation of the effect sizes from the linear mixed effects model showed the threshold increase with increasing intraocular straylight to be consistently smaller with the AC stimulus than the threshold increase observed with other stimulus forms, when considering all 18 test locations, and on evaluation of individual eccentricity zones. Threshold increase from baseline was observed to be largely similar with increasing intraocular straylight in all five eccentricity zones, although there were some differences in the fog filter with which a statistically significant increase in threshold was first noted. Threshold increase from baseline with increased straylight reached statistical significance with similar levels of intraocular straylight in all five eccentricity zones for the A and AC stimuli. For the C_R and GIII stimuli, threshold increase from baseline reached statistical significance with systematically lower levels of intraocular straylight with increasing eccentricity from fixation, within the central 21.2° (i.e. zones

1-3). Beyond 21.2°, threshold increase from baseline was only statistically significant with greater levels of intraocular straylight.

6.5   Discussion

It is important to consider the purpose of perimetric testing when considering the vulnerability of a stimulus to optical imperfections. A perimetric test is conducted to identify changes in visual function associated with neural tissue damage or loss. As discussed in section *6.1*, optical imperfections such as blur and straylight interfere with the focus of light on the retina; more light energy is then required to elicit a cellular response, this being measured as a reduction in contrast sensitivity. This may mimic the reduction in sensitivity that is noted as a result of neural damage, such that it becomes difficult to distinguish neural damage from optical interference. An ideal perimetric stimulus would remain robust to the effects of optical imperfections, ensuring that a reduction in perimetric sensitivity could reliably be attributed to a reduced neural response.

Two potential sources of optical interference have been investigated here, both of which are common occurrences in clinical practice. Inadequate refractive error correction, human error, and off-axis refractive error, mean that optical defocus will occur at some, if not all, test locations in every patient undertaking a perimetric test. The age demographic of patients typically undertaking a perimetric test is such that aging changes in the intraocular lens, in addition to more clinically significant cataract, are highly prevalent, resulting in an increase in intraocular straylight.

Although both types of optical imperfection may result in degradation of the retinal image, the robustness of a perimetric stimulus to one type of optical imperfection may not predict the robustness of the same perimetric stimulus to optical imperfections of a different nature. This has been demonstrated in FDT, which appears to be largely robust to the effects of optical defocus (Anderson and Johnson 2003; Artes et al. 2003b; Horner et al. 2013), yet shows a greater vulnerability to intraocular straylight than SAP (Anderson et al. 2009; Bergin et al. 2011). It is therefore necessary to separately quantify the effects of optical imperfections with any novel perimetric stimulus, as demonstrated here.

6.5.1　Optical defocus

In the examination of optical defocus induced by positive, spherical lenses, threshold increase from baseline was statistically significant with both +2.00 DS and +4.00 DS lenses with all four stimulus forms. Comparing threshold differences from baseline in *Figure 6.4.B*, small differences may be observed between stimulus forms, with the C_R stimulus showing the greatest increase in threshold from baseline with both +2.00 DS and +4.00 DS. While this corresponds with previous studies, which noted a greater vulnerability to optical defocus with smaller stimuli (Sloan 1961; Atchison 1987; Anderson et al. 2001), no statistically significant difference was found between stimulus forms when analysed with a linear mixed effects model.

It would be reasonable to expect a greater response variability with an increase in optical defocus, however this does not appear to be the case. No statistically significant difference in response variability with increasing optical defocus was noted with the A, C_R, or GIII stimuli. Although a statistically significant difference was observed in response variability with increasing optical defocus with the AC stimulus, the response variability was, in fact, lower with both +2.00 DS and +4.00 DS in comparison with that at baseline. No particular trend was observed; response variability with +4.00 DS was lower than that at baseline, but higher than that with +2.00 DS. This is, therefore, unlikely to be of clinical importance, and likely represents random variation, particularly given that no statistically significant difference in response variability was noted between stimulus forms.

Although not directly tested here, one might be tempted to conclude that an increase in other optical imperfections will also have a limited impact on response variability. However, as previously indicated, caution should be exercised when attempting to predict the effects of one type of optical imperfection from the effects of another.

As noted in chapter four, it is important to consider both signal (in this experiment, the increase in threshold with increased optical defocus) and noise (response variability) together, in addition to considering them separately, as was carried out in the SNR analysis. Contrary to the experiment presented in chapter four, in which a higher SNR indicated an improved detection of glaucomatous from normal test locations (i.e. a

desired effect), a higher SNR in this experiment indicates a greater vulnerability of the stimulus to the effects of optical defocus (i.e. an unwanted effect). The GIII stimulus appeared to indicate a substantially lower SNR than the other three stimulus forms in all conditions of optical defocus (*Figure 6.6.A*). However, it is more appropriate to consider the increase in SNR from baseline to understand the effects of optical defocus itself; this was still lower with the GIII stimulus in comparison to the other three stimuli, but was less pronounced (*Figure 6.6.B*), and no statistically significant differences were found between stimulus forms.

To calculate TD, the estimated threshold for each participant's age was extrapolated from the linear regression model of threshold against age from section *4.3* (*Figure 6.2*). Although the relationship between age and threshold was linear for each of the four stimulus forms for the age range tested in chapter four (51.0-86.2 years), it is possible that this relationship is not strictly linear when considered over a wider age range, which may lead to anomalies in the SNR calculation for younger participants. While threshold measures for younger participants with area-modulating stimuli, such as the A and AC stimulus forms, are as yet unknown, various studies have investigated the effects of age on threshold with contrast-modulating stimuli. Heijl et al. (1987) investigated perimetric sensitivity in 95 participants between 20 and 80 years of age. While their findings appear to indicate a linear relationship between age and threshold across this age range, there appears to be some heteroskedastic characteristics in this relationship; Figure 1 in Heijl et al. (1987) indicate a greater range of perimetric sensitivity values in participants aged 20-30 years, compared with older participants. This observation of greater inter-observer variability in younger participants with a Goldmann III stimulus may explain the substantially lower TD observed with the GIII stimulus, and as such a substantially lower SNR, compared with other stimulus forms in this experiment. It is possible that a lower inter-observer variability exists with the A, AC, and C_R stimuli with younger participants compared with a Goldmann III stimulus, which may explain why TD with these stimulus forms were closer to that predicted by the linear regression model. Most studies examining the effects of age on smaller stimuli have done so using kinetic perimetry, e.g. Grobbel et al. (2016); this shows how isopter eccentricity changes with age, but does not provide quantification of threshold

changes with age as investigated here. It is therefore difficult to predict the relationship between age and threshold for the C_R stimulus from existing studies, although the findings of this experiment indicate a consistency with that of the A and AC stimuli, rather than the GIII stimulus. This may be due to the C_R stimulus undergoing complete spatial summation, while the GIII stimulus underwent incomplete spatial summation, at the locations tested in this experiment.

Overall, the findings of this experiment suggest little clinical difference in robustness to optical defocus between all four stimulus forms. As in the examination of test-retest variability described in chapter five, further investigation is now recommended with the application of a clinical thresholding algorithm, particularly for the A stimulus, in which a greater SNR than that of a Goldmann III equivalent stimulus was demonstrated (chapter four). It is possible that some differences between stimulus forms could be observed with finer increments of induced optical defocus from baseline (e.g. +0.50 DS to +2.00 DS), rather than the more gross steps used in this experiment. Further investigation with the application of a clinical thresholding algorithm would therefore benefit from inclusion of finer blur increments than used here.

In addition, it would be beneficial to investigate the effects of optical defocus at test locations with a higher baseline threshold, i.e. a larger stimulus area at threshold with the area-modulating stimuli. Previous studies have demonstrated an increased robustness to the effects of optical defocus with larger, fixed-area stimuli (Sloan 1961; Atchison 1987; Anderson et al. 2001). This could be achieved by testing participants with glaucoma in future studies, using a similar experimental design as utilised here.

It may also be advisable to attempt to investigate differences in threshold with optical defocus across various eccentricities, although as previously discussed this is unlikely to be a straightforward task. Given the findings of studies examining peripheral refractive error (Ferree et al. 1931; Ferree and Rand 1933; Rempt et al. 1971; Millodot 1981; Atchison et al. 2006), it is possible that some observers may experience an improvement in threshold at certain peripheral test locations with the introduction of positive spherical lenses, as these lenses may, in fact, correct the refractive error at

that location, while experiencing a decline in threshold at other test locations. It is therefore likely that a greater variability in inter-observer thresholds will be observed.

6.5.2 Straylight

6.5.2.1 Repeatability of C-Quant straylight values

In evaluating the repeatability of the C-Quant straylight meter, the findings presented here were largely in agreement with those of previous studies, which have indicated highly reproducible straylight measurements (Franssen et al. 2006; Cervino et al. 2008; Guber et al. 2011). It is, however, worth noting that many of the baseline measurements presented here were higher than average, and even higher than the 95% confidence limits, according to the instrument's normative database. Although a healthy participant may demonstrate above-average straylight in an experiment such as this, a higher proportion was found than may be expected. Given that there were no statistically significant differences between participants, this cannot be attributed to an anomalous observer; equally, there were no statistically significant differences between the three sets of readings, such that this may not be attributed to high intra-observer variation. Furthermore, all measurements were indicated as reliable according to the instrument's reliability indices.

The manufacturer's guidelines warn that squeezing of eyelids, and eye position in relation to the instrument, may influence measurements. Although care was taken to follow these guidelines, it is possible that these factors may have influenced results, particularly as most participants in this experiment had no prior experience with the C-Quant straylight meter. However, one participant had prior experience with the C-Quant; this participant's mean straylight measures are denoted by the ▲ symbol in *Figure 6.11*, which are also observed to be above the 95% confidence limits at baseline. It is therefore unlikely that these findings are due to observer inexperience.

Manufacturer's guidelines state that refractive correction is not always necessary for use with the C-Quant, and the use of a single, mean sphere lens is recommended. In this experiment, full refractive correction was worn by all participants, as experimental tests with the four stimulus forms were also conducted in this manner; it was therefore necessary to measure the straylight values under the same refractive

conditions. Care was taken to ensure that all lenses used were clean, to avoid anomalous results, but it is possible that the use of these lenses increased the resulting measurements, such that log straylight values were higher than those included in the normative database. However, it is worth noting that one participant did not require any refractive correction; this participant's straylight measures are denoted by the ▼ symbol in *Figure 6.11*, which are also above the 95% confidence limits at baseline.

The straylight measurements used in the normative database of the C-Quant straylight meter are taken from a study by Van den Berg (2007), examining change in log straylight with age. Figure 1 from this study shows the measurements used in the C-Quant's normative database, from which several straylight measurements consistent with the findings presented here, can be observed in subjects similar in age to those included in this experiment. It is worth noting that there are fewer, younger subjects (< 45 years) included in this normative database, compared with older subjects. Therefore, the 95% confidence limits included in the C-Quant's normative database may not be truly representative of this age-group.

Straylight values appear more variable at baseline than with the fog filters; this cannot be attributed to test order, as participants completed tests for all straylight conditions in random order. The increase in threshold from baseline (expressed as a percentage) with the fog filters also does not appear consistent with that of Bergin et al. (2011), who reported an increase in intraocular straylight of 10% to 200% with the same five fog filters. However, it is not clear whether the percentages quoted by Bergin et al. (2011) represent the mean straylight increase for their six participants, or the full range (i.e. the minimum percentage increase with Fog 1, and the maximum percentage increase with Fog 5). In the experiment presented here, the minimum percentage increase with Fog 1 was 8.0%, and the maximum percentage increase with Fog 5 was 110.0%, which is more consistent with that of Bergin et al. (2011), although still shows a lower increase in straylight with Fog 5.

As previously discussed in section *6.1.2*, there are indications in the literature that iris pigmentation (Van den Berg et al. 1991; Van den Berg 1995), and refractive error status (Rozema et al. 2010), may influence straylight measures. The findings presented

here are in general agreement with those reported in previous studies. Log straylight values with myopic participants were higher than with the hyperopic participant, as found by Rozema et al. (2010), although it was not appropriate to test this statistically in this experiment. Although not statistically significant, straylight values were found to be higher in those participants with lightly pigmented irides, in comparison to those with darker irides. Although some studies have noted a statistically significant increase in straylight with lighter-coloured irides (Van den Berg et al. 1991; Van den Berg 1995), other studies have noted an increase in straylight, but the difference was insufficient to achieve statistical significance, as observed in this experiment (Rozema et al. 2010; Guber et al. 2011). The findings presented here may in part be influenced by the categorisation of hazel irides as 'light pigmentation'. Although hazel eyes contain less pigmentation than brown eyes, they do contain more pigmentation than blue eyes. As such, it is possible that less of a difference is noted between the two pigment groups here, than if all three of these participants had blue eyes.

6.5.2.2 Effect of straylight

Threshold increase with increasing straylight appeared curvilinear with the A and AC stimuli, although perhaps not as originally hypothesised. As noted in section *6.2*, it was thought that the initial increase in threshold due to straylight would be steeper, but would slow once threshold was beyond a certain area. Contrary to this hypothesis, initial increase in threshold due to straylight was slower, with a steeper increase demonstrated with denser fog filters. Although a similar increase in threshold was noted with the GIII and C_R stimuli with increasing straylight, threshold increase appeared slightly more linear with these stimuli. Mean increase in straylight was 0.83 log(s) from baseline to Fog 5, and mean increase in threshold with the GIII stimulus was 0.40 cd/m^2.s.deg^2 from baseline to Fog 5. Converted to HFA-equivalent dB, this represents a decrease in SAP sensitivity of 4.01 dB, which is a greater decrease than reported by Bergin et al. (2011), but is consistent with the findings of Anderson et al. (2009).

In examining threshold increase with increased intraocular straylight, no statistically significant difference is found between stimulus forms when directly compared, either when all 18 test locations are considered, or when individual eccentricity zones are

considered. However, by analysing each stimulus form separately, and determining the fog filter at which threshold increase first becomes statistically significant, modest differences can be observed between stimuli.

When considering all 18 test locations, the AC stimulus appeared the most robust to straylight, as threshold increase was consistently lower than other stimulus forms with all fog filters, and a statistically significant increase in threshold was not observed until Fog 4, consistent with clinically significant cataract, in accordance with the three categories of log straylight defined by Bergin et al. (2011). In comparison, a statistically significant increase in threshold is observed with Fog 2 with the A and C_R stimuli. Threshold increase with increasing intraocular straylight was observed to be highest with the GIII stimulus for fog filters 1-3 (consistent with normal aging changes), with a statistically significant increase in threshold with all fog filters.

On examination of the different eccentricity zones, threshold increase with increased straylight was again generally consistently lower with the AC stimulus in comparison with the other stimulus forms. Although the stimulus which demonstrated the greatest threshold increase varied between eccentricity zones, and between straylight condition, the GIII stimulus often demonstrated the greatest threshold increase with fog filters 1-3, consistent with normal aging changes. Some anomalous results were noted in zone 3 for the A stimulus, as a statistically significant increase in threshold is observed with Fog 2, but not with Fog 3; increase in threshold is statistically significant again with Fog 4. It is likely the statistical significance with Fog 2 represents a type I error, despite the post hoc correction, given the lack of significance with the subsequent fog filter.

The robustness of the C_R and GIII stimuli to intraocular straylight appeared to be dependent on eccentricity, with increase in threshold reaching statistical significance with systematically lower levels of intraocular straylight within the central 21.2°, but increase in threshold from baseline only becoming statistically significant with higher levels of intraocular straylight beyond this eccentricity. However, observation of the effect sizes showed threshold increase with increasing intraocular straylight to be largely similar between eccentricity zones for all four stimulus forms.

There is potentially more intraocular straylight with a stimulus of larger area in comparison to that of a smaller area. This may explain the higher threshold with the GIII stimulus in zones 1 and 2 (*Figure 6.13, A & B*), as thresholds with the A, AC, and C_R stimuli all had a smaller area at threshold. Given that the A and AC stimuli do not present a fixed stimulus area, but modulate this throughout the test to determine threshold, there is the possibility that the effects of straylight will differ with differing baseline thresholds, as hypothesised for the effects of optical defocus in section *6.5.1*.

To aid in this investigation, the threshold at baseline within each of the five eccentricity zones can be examined. If a greater amount of straylight was observed with larger stimuli, one might expect a notably higher baseline threshold with the largest stimulus at threshold in each of the five eccentricity zones. As previously noted, the GIII stimulus does demonstrate a higher threshold within zones 1 and 2, those closest to fixation where it is the largest stimulus at threshold. However, as observed in *Figure 6.13.E*, baseline threshold with the A stimulus in zone 5 (mean 0.12 deg^2, the largest area at threshold in this zone), does not appear substantially greater than that with the C_R stimulus (0.01 deg^2), despite the difference in stimulus area. If straylight reduces contrast by a fixed, linear amount, a greater reduction in total energy would be observed with a larger, dimmer stimulus (e.g. the A stimulus at a higher baseline threshold), than a smaller, brighter one. Again however, observation of *Figure 6.14.E*, which shows the difference in threshold with increase in straylight, does not indicate a substantially greater increase in threshold with the larger, A, stimulus compared with that of the substantially smaller C_R. Peripheral refractive error may be a confounding factor in this observation, which may not permit firm conclusions. Further examination with participants of differing threshold, for example those with glaucoma, may permit a closer investigation of the effects of straylight on an area-modulating stimulus, although as this would involve testing older participants, great care would be required to ensure that existing intraocular straylight due to aging changes was not a confounding factor.

As noted in chapter five, the use of a flat OLED display may be a limitation of this experiment, as stimuli will be subject to some distortions of shape and luminance due to viewing angle. As in chapter five, however, as the raw threshold measurements

were not the primary focus of this experiment, but the difference in threshold due to the introduction of additional straylight, this should not impact on the conclusions drawn here; test locations remain constant, and as all participants completed tests for each stimulus form, they served as their own control.

As the fog filters used here also cause a reduction in luminance with angular distance from a light source (Zlatkova et al. 2006), it is perhaps surprising that no ceiling effect was observed. In particular, one might expect a ceiling effect to be observed with the C_R stimulus, as this achieved the lowest dynamic range of the four stimulus forms with the apparatus used, given the extent of luminance reduction noted within the central 15° alone by Zlatkova et al. (2006). It is also worth noting that the straylight values measured with the C-Quant are those for central vision, and this may be different for those peripheral locations tested here; further investigation may benefit from attempts to quantify peripheral straylight values.

Overall, the findings of this experiment suggest that the AC stimulus may demonstrate an increased robustness to intraocular straylight compared with the other stimulus forms. However, as no statistically significant difference was found when comparing stimulus forms directly, this likely indicates that any differences in straylight robustness between stimulus forms is modest at best. Anderson et al. (2009) noted that, when considering the 95% confidence limits of the normative database for each of the perimetric instruments used, only a minority of data points fell outside this range. As yet, it is not possible to examine this with the stimuli presented here, as 95% confidence limits for the A, AC and C_R stimuli have not been established, and in the absence of known 95% confidence limits, it is difficult to ascertain the clinical significance of any noted differences. As in the investigation of optical defocus, further investigation is now recommended with the application of a clinical thresholding algorithm, particularly for the A stimulus, in which a greater SNR than that of a Goldmann III equivalent stimulus has been demonstrated (chapter four). Establishment of 95% confidence limits would permit a firmer conclusion to be drawn as to the clinical significance of any findings, and testing with a greater number of visual field locations would aid in the examination of straylight impact on eccentricity.

Chapter 7 Overall discussion and future work

7.1   Overall discussion

The work presented in this thesis, investigating the optimisation of perimetric stimuli for mapping changes in spatial summation in glaucoma, tests the hypothesis of Redmond et al. (2010a) that a stimulus modulating in area, either instead of or simultaneously with contrast, would better identify the lateral shift in the spatial summation curve in glaucoma. Stimuli have been optimised to exploit this change in spatial summation, in an attempt to boost the glaucoma disease signal, with the proposal that this would additionally reduce both intra- and inter- test variability.

Stimulus areas used here have not been restricted to those established by Goldmann (Goldmann 1945a; Goldmann 1999), which were originally developed for use with his manual, kinetic perimeter. As discussed in chapter one, Goldmann stimuli were designed for use with complementary luminance values, such that equivalent isopters could be plotted by changing stimulus area or stimulus luminance accordingly. Given that automated perimetry is largely concerned with static thresholding tests, the continued use of Goldmann stimuli may be considered historical baggage from kinetic testing. The dB scale also originated from kinetic, manual perimetry, and in applying a measure of perimetric sensitivity relative to the luminance capabilities of the instrument, hinders the comparison of sensitivity measures from one instrument to another. Perimetric research would benefit from thinking beyond these restrictions.

The stimulus forms utilised in this series of experiments comprised of one stimulus modulating in area only (A) while maintaining a constant contrast, designed such that the area modulations began within the area of complete spatial summation. One stimulus modulated in both area and contrast simultaneously (AC), from a small, dim stimulus, to a larger, brighter stimulus, again designed such that the modulations began within the area of complete spatial summation. One stimulus modulated in contrast only, and was smaller than the currently utilised Goldmann III stimulus, such that it largely operated within the area of complete spatial summation at all locations of the visual field (C_R). Increment steps, whether area, contrast, or both, were matched between these three stimulus forms, ensuring that the raw energy in each

step, determined in accordance with *Equation 3.3*, were equivalent, irrespective of stimulus type. One additional stimulus was equivalent to that of a standard Goldmann III stimulus, which modulated in contrast only (GIII), and thus served as a control stimulus. This stimulus used the same contrast scale as the C_R stimulus. All findings were presented in log energy (cd/m^2.s.deg^2), enabling direct comparisons to be made between stimulus forms.

As detailed in chapter two, a commercially available projection perimeter, as is commonly used in SAP testing, could not accommodate the required area increments. As such, the ViSaGe MKII visual stimulus generator (Cambridge Research Systems, Rochester, UK), in combination with a Sony PVM-A250 Trimaster El OLED display, were selected to run the experiments in which the four stimulus forms were compared. This permitted greater control and flexibility over stimulus generation and display.

As discussed in chapter three, previous studies have reported that the use of a smaller perimetric stimulus area resulted in greater variability of the sensitivity, while the use of a larger perimetric stimulus area resulted in lower variability of the sensitivity (Wall et al. 1997; Wall et al. 2013). As the series of experiments presented in this thesis involved the use of small stimuli, and stimuli of non-constant area, this was an important factor to consider. An examination of the work of Hill (2001) highlighted the need for careful pre-selection of stimulus values for use with a MOCS procedure, and further scrutiny of the study of Wall et al. (1997) indicated some aspects of the methodology which may have introduced unintentional bias. As such, it was deemed prudent to further investigate the response variability characteristics of stimuli of differing areas. This was investigated in five young, healthy participants, using standard Goldmann stimuli I-V, and an Octopus 900 perimeter.

By adopting a two-phase protocol, informed by the studies of Hill (2001), an approximate shape of the FOS curve could be established, and thus more appropriate pre-selection of stimulus values could be undertaken before conducting a standard MOCS procedure. By this method, and by not restricting stimulus selection to integer dB values, approximately equal numbers of sub- and supra- threshold stimulus

presentations could be made with all five Goldmann stimuli, permitting a more direct comparison to be made between response variabilities with different stimulus areas.

This experiment demonstrated similar response variability characteristics between the five Goldmann stimuli, which did not support previous findings of a reduced response variability with larger stimuli. Although this experiment did not directly test those with glaucoma, areas of the visual field with a low sensitivity were utilised, with no obvious difference in response variability observed between Goldmann stimuli.

While further testing should be conducted in those with established glaucoma, equating raw energy steps across differing stimuli, and comparing SNR, these findings call into question the proposed shift from Goldmann III to Goldmann V in those with glaucomatous damage (Wall et al. 1997; Wall et al. 2013; Gardiner et al. 2015), which has largely been based on variability characteristics. Findings also indicated that the proposed area-modulating stimuli used in this series of experiments should not be subject to a non-constant response variability due to stimulus area.

In the experiment presented in chapter four, 20 healthy participants, and 30 participants with glaucoma, underwent testing with each of the four stimulus forms, three optimised to probe spatial summation changes that occur in glaucoma, and one equivalent to a Goldmann III stimulus, as is commonly used in SAP, as a control stimulus. A MOCS procedure, using a three-phase protocol (the two-phase protocol from chapter three with an additional, initial staircase phase to establish an estimated threshold), was utilised. Four, equidistant test locations were used, to control for attentional bias.

All three stimulus forms optimised to probe spatial summation changes (A, AC, and C_R) demonstrated a higher disease signal than that of the GIII stimulus, with the A stimulus showing the greatest overall disease signal, supporting the hypothesis of Redmond et al. (2010a).

Comparisons of response variability across three disease strata demonstrated that response variability was most dependent on depth of defect with the Goldmann III equivalent (GIII) stimulus, as response variability increased the most with disease

severity with this stimulus. This was in keeping with previous studies (Wall et al. 1996; Henson et al. 2000), which have demonstrated a substantial increase in response variability with depth of defect with a Goldmann III stimulus. Response variability was found to be least dependent on depth of defect with the A and C_R stimulus forms.

As neither disease signal, nor response variability alone can fully determine the advantage of one stimulus form over another, analysis of SNR was perhaps the most crucial comparison made between the four stimulus forms. Here, all three stimulus forms optimised to probe spatial summation changes (A, AC, and C_R) demonstrated a higher SNR than the GIII stimulus. The A stimulus in particular demonstrated a distinct advantage over the GIII stimulus, with the difference in SNR between the A and GIII stimuli noted to systematically enlarge with increasing disease severity. Although this was not a longitudinal study, and therefore identification of disease progression has not yet been investigated, this finding holds promise that the A stimulus might also outperform the GIII stimulus in this regard.

Chapter five details an experiment conducted into the test-retest characteristics of each of the four stimulus forms. Test-retest variability is a well-documented limitation of SAP, and has been shown to enlarge with increasing disease severity (Wilensky and Joondeph 1984; Heijl et al. 1989a; Artes et al. 2002a). As such, it is an important consideration in any new stimulus paradigm, as test-retest characteristics are necessary in establishing the accuracy with which early glaucomatous damage can be distinguished from normal, and disease progression can be identified from stable disease.

A 1:1 adaptive staircase strategy with four reversals was used to test 18 visual field locations, informed by the findings of Wang and Henson (2013), at a range of eccentricities from fixation. Fifteen participants with glaucoma, with a range of disease severity, and five healthy participants were tested, each completing five tests for each of the four stimulus forms. By including participants who had prior experience of the A, AC, and C_R stimuli, in addition to participants who only had experience with a standard Goldmann III stimulus, an analysis could also be conducted into the learning and fatigue effects of each stimulus form.

Comparisons of test-retest characteristics were made by determining the area between the 5th and 95th retest percentiles for each stimulus form, and separate analyses were conducted with test locations that underwent complete spatial summation with all stimulus forms, and those that underwent incomplete spatial summation with a standard Goldmann III.

Overall, the test-retest characteristics of each stimulus form appeared largely similar, with no one stimulus form demonstrating substantially lower test-retest variability. In addition, spatial summation characteristics did not appear to influence test-retest variability.

The well-documented increase in test-retest variability with eccentricity from fixation (Katz and Sommer 1986; Heijl et al. 1987; Heijl et al. 1989a; Heijl et al. 1989b; Chauhan and House 1991) could have been attributed to a fatigue effect, as the testing strategies used in these studies initially establish perimetric sensitivity at four seed locations near fixation, and then subsequent stimuli are generally presented more peripherally as the test progresses. However, as the findings presented in chapter five are similar to those of Artes et al. (2002a), despite stimulus presentation occurring at truly randomised test locations, this experiment confirms that variability with eccentricity from fixation cannot be attributed to a fatigue effect alone.

A previous study by Gardiner et al. (2014) hypothesised that increased test-retest variability at areas of low sensitivity were due to saturation of the retinal ganglion cells. Given that test-retest variability observed in this study is similar between all stimulus forms, including the A stimulus in which contrast is a low, constant value, this supports the findings of Anderson et al. (2016), and provides further evidence that ganglion cell saturation is not the cause for the increased test-retest variability in areas of low sensitivity.

No additional learning or fatigue effect was noted with the novel stimulus forms (A, AC, C$_R$), although as all participants had experience with a standard Goldmann III stimulus, it is possible that those who are completely inexperienced with perimetric testing may exhibit a fatigue/learning effect. It is hypothesised that participants who are perimetrically inexperienced would display similar learning characteristics to those

found with a Goldmann III in SAP, in that a minority of observers would display a significant learning effect, but the majority of observers would produce reliable results with little practise (Heijl et al. 1989b).

The importance of successful distinction between neural loss, as occurs in the presence of glaucoma, and apparent sensitivity loss due to optical imperfections, is discussed in chapter six. Differences in threshold as a result of two types of optical imperfection are investigated with the four stimulus forms; those differences that occur due to optical defocus, and those that occur due to intraocular straylight. These optical imperfections were selected as they are common occurrences in clinical practice.

Several types of perimetric stimulus are demonstrably vulnerable to optical imperfections, and it is important to note that, despite all optical imperfections resulting in a degradation of the retinal image, a robustness to one type of optical imperfection is not a predictor of robustness to other optical imperfections (Herse et al. 1998; Kim et al. 2001; Anderson and Johnson 2003; Anderson et al. 2009; Bergin et al. 2011). Therefore, this must be characterised with any new stimulus paradigm. Given that smaller stimuli are reportedly more vulnerable to optical imperfections than larger stimuli (Sloan 1961; Campbell and Green 1965; Radius 1978; Atchison 1987; Anderson et al. 2001; Horner et al. 2013), an informed hypothesis could be made with respect to the relative effect of optical imperfections on threshold with the C_R and GIII stimuli. However, it was much more difficult to predict their effects on the A and AC stimuli, in which area is not constant.

In the examination of optical defocus, as induced by positive, spherical lenses, the same three-phase MOCS procedure was used as that described in chapter four, to determine the effects on both threshold and response variability at test locations equidistant from fixation.

Similar findings were found with all four stimulus forms, with the addition of a +2.00 DS lens found to induce a statistically significant increase in threshold with all stimulus forms. Although the C_R stimulus demonstrated the greatest increase in threshold with

increasing optical defocus, no statistically significant difference was found in threshold increase between stimulus forms.

Although an increase in response variability may be expected with an increase in optical defocus, this was not demonstrated in this experiment. No statistically significant increase in response variability was found with increasing optical defocus with any of the four stimulus forms.

Both threshold and response variability were considered together, as SNR. Although SNR increased with increasing optical defocus with all four stimulus forms, no statistically significant difference was found in SNR increase between stimuli.

Five white, opacity containing filters (Fog 1-Fog 5 Standard; LEE Filters, Andover, UK) were used to induce additional straylight. Five young, healthy participants undertook the same test as used in chapter five, with threshold established at 18 test locations using a 1:1 adaptive staircase strategy of four reversals, under each of the six straylight conditions, with each of the four stimulus forms. Threshold increase with increasing straylight appeared curvilinear with the A and AC stimulus forms, but slightly more linear with the C_R and GIII stimulus forms. Findings for the GIII stimulus were found to be in keeping with those reported with SAP by Anderson et al. (2009), who used the same fog filters as used here.

The AC stimulus appeared most robust to the effects of intraocular straylight, as threshold increase with increasing intraocular straylight was consistently lower with this stimulus in comparison with other stimulus forms. When considering all 18 test locations, a statistically significant increase in threshold was only observed from baseline with Fog 4 and Fog 5, those that might be considered to be aligned with clinically significant cataract. The GIII stimulus appeared the least robust to intraocular straylight when considering all test locations, as the addition of all fog filters resulted in a statistically significant increase in threshold from baseline (Fog 1-Fog 3 being consistent with normal lenticular aging changes).

When considering different zones of eccentricity, findings with the A and AC stimulus forms appeared consistent across all five zones, as threshold increase from baseline

was generally only statistically significant with Fog 4 or Fog 5 in all zones. Findings for the C_R and GIII stimulus forms appeared to be more dependent on eccentricity from fixation. However, observation of the effect sizes showed threshold increase with increasing intraocular straylight to be largely similar between eccentricity zones for all four stimulus forms.

While these findings indicate that area-modulating stimuli, and particularly the AC stimulus, may be more robust to the effects of straylight, a direct comparison did not indicate any statistically significant difference between stimulus forms.

7.2 Overall conclusions

Overall, the findings of this study suggest that a stimulus modulating in area only, with no modulation in contrast, may offer benefits for measuring glaucomatous changes in spatial summation in a clinical setting, in the form of a greater disease signal, a more uniform response variability with depth of defect, and a greater SNR, when compared with a standard, contrast-modulating Goldmann III stimulus of fixed area.

Test-retest variability appears to be largely similar with a stimulus modulating in area only, in comparison with a standard Goldmann III stimulus, as is its robustness to optical defocus. There is some indication that an area-modulating stimulus may be more robust to the effects of intraocular straylight when compared with a contrast-modulating Goldmann III stimulus, although further investigation is recommended as detailed in chapter six. While a stimulus modulating in both area and contrast may show a further increase in robustness to the effects of intraocular straylight, this stimulus did not demonstrate as great a difference in SNR in comparison with a Goldmann III equivalent stimulus, as was demonstrated with a stimulus modulating in area only.

7.3 Limitations of this study

It is important to recognise that the dynamic ranges of each of the four stimulus forms are specific to the hardware used. In a Humphrey Field Analyzer or Octopus projection perimeter, the maximum luminance output available is 3183.1 cd/m^2, giving a maximum log energy of 1.97 cd/m^2.s.deg^2 for a Goldmann III stimulus. In comparison,

if using a standard test grid, the maximum diameter of any stimulus would be 6°, to avoid overlapping with adjacent test locations, giving a maximum log energy of 1.45 cd/m^2.s.deg^2 for an A stimulus. Any stimulus modulating in both area and contrast should give the largest dynamic range in a perimetric instrument, if not limited by area or contrast. In comparison, in the series of experiments presented here, the A stimulus had a higher maximum energy than the GIII stimulus, given the luminance restrictions of the OLED display.

However, for the purposes of a study such as this, it is important to compare stimulus forms on the same platform. As noted in chapter two, this was not possible on a commercially available projection perimeter, due to the restrictions in available stimulus area. By using the same platform for each stimulus form, and by using equivalent starting energy values, approximately equal increment step-sizes in the energy of the A, AC and C$_R$ stimulus forms could be used, ensuring these stimuli were operating at the same position of the logarithmic energy scale. This was important in ensuring that stimulus forms could be directly compared. Given the importance of using the same platform for all stimulus forms, the hardware chosen here was the most suitable for this series of experiments. Although restricted in luminance output, it allowed for much finer increments of area to be used.

The lower maximum luminance output of an OLED display could be seen as a limitation, however was considered appropriate given that participants were viewing the light source directly. If an OLED display incorporated a much higher maximum luminance output, similar to that of a projection perimeter, after-images could have been induced subsequent to a bright stimulus presentation; this is less of a concern with a projection perimeter, as observers do not view the light source directly. Such after-images could have obscured subsequent stimulus presentations, and therefore influence the findings of the study. Therefore, the study set-up and apparatus used were optimal for this series of experiments, and no such after-images are expected to have occurred with the maximum luminance output of the Sony PVM-A250 Trimaster El OLED display.

This is an important consideration in future work, if taking an area-modulating stimulus into a clinical setting. The area increments on a projection perimeter are denoted by those available in the incorporated aperture wheel, which currently only include Goldmann I-V stimulus areas. Whether a different aperture wheel, with a wider range of area increments, used in a projection perimeter, or whether a monitor-based instrument, such as is used in the Humphrey FDT/Matrix, or the Heidelberg Edge Perimeter (HEP), will require careful consideration to ensure the chosen instrument is optimal for a clinical setting. If a flat, monitor-based instrument is selected, the previously highlighted limitations of shape and luminance distortion of presented stimuli at the periphery of the display screen (chapters five and six), should be addressed.

7.4   Future work

While the findings detailed in this thesis demonstrate the promise of an area-modulating stimulus, these experiments only represent the initial stages of investigation, and further work is required to determine whether this stimulus would present a significant advantage in a clinical setting, above the currently used Goldmann III stimulus.

7.4.1   Thresholding algorithm

As yet, all experiments have been conducted using basic, psychophysical techniques, thus ensuring that findings were due to the stimulus form itself, and not due to confounding factors. In particular, the MOCS procedure permitted investigation of response variability, and minimised much of the bias that may occur with other testing strategies. A clinical thresholding algorithm should now be applied, to permit a greater number of visual field locations to be tested in an acceptable time-period, and thus permit investigation of the proposed area-modulating stimulus within a clinical setting.

As current thresholding algorithms, such as SITA (Bengtsson et al. 1997a), ZEST (King-Smith et al. 1994), and QUEST (Watson and Pelli 1983), were all developed for use with a standard Goldmann III stimulus, it is necessary to develop a new clinical thresholding algorithm, specifically designed for use with an area-modulating stimulus. This is the next logical step in the development of a test incorporating this stimulus. Subsequent

investigations into the clinical utility of this stimulus should be performed once an appropriate thresholding algorithm has been developed.

7.4.2   Test-grid

Studies have demonstrated that the existing patterns of test locations, such as the 24-2 and 30-2 test-grids, are not evenly distributed with respect to retinal ganglion cell density (Garway-Heath et al. 2000a), and it has been suggested that the use of test locations that are more representative of this distribution may improve correlations between retinal structure and visual function (Asaoka et al. 2012; Malik et al. 2012). However, as increasing evidence indicates that Ricco's area is likely a perceptual result of spatial filtering at multiple hierarchies of visual processing, in the retina and at the visual cortex (Ransom-Hogg and Spillmann 1980; Schefrin et al. 1998; Vassilev et al. 2000; Anderson 2006; Pan and Swanson 2006; Je et al. 2018), i.e. the 'perceptive field' (Vassilev et al. 2005; Anderson 2006), a test-grid designed to maximise distribution of retinal cells may not be the most optimal. Investigation into the most optimal test-grid is recommended, and it is likely this will result in test locations that are closer together in the central visual field, and further apart in the peripheral visual field. As such, a differing maximum area may be beneficial at differing eccentricities to avoid overlapping adjacent test locations. Additionally, investigation of the optimal test-grid should not be restricted to the central 30°, as benefit may be found in a wider test area.

7.4.3   Multi-centre trial

Once a thresholding algorithm and optimal test-grid have been established, the next logical step would be a large, multi-centre trial. This would permit testing with a larger number of participants from a wider population, and a comparison with the current clinical standard. Testing of both healthy participants, and those with a wide range of glaucomatous severity, would enable a thorough investigation to establish the clinical utility of an area-modulating stimulus. Ideally, three stages would be undertaken, as follows.

7.4.3.1 Cross-sectional study

The cross-sectional study proposed here would utilise the area-modulating stimulus with its optimal thresholding algorithm and test-grid, and would compare findings with clinical SAP. By undertaking this study as part of a multi-centre trial, it would be possible to test a much greater number of participants, both healthy and those with glaucoma. Care would be required to ensure that appropriate comparisons were made between the two strategies, given the differences between algorithms, test grids, and measurement scales. This may be achieved with some of the methods discussed in section *1.3.4*, such as SNR (Artes and Chauhan 2009) or PoPLR (Redmond et al. 2013a), used to compare between SAP and FDT, which also incorporate substantially different test strategies. Testing a large cohort of healthy participants, of a wide range of age-groups, would permit collation of data into a normative database, for future use in distinguishing a normal visual field location from early glaucoma defects. It would be beneficial to ensure that similar numbers of participants are included in each age group.

7.4.3.2 Test-retest variability

As discussed in chapter five, test-retest characteristics may be influenced by the use of a thresholding algorithm. As such, once a suitable algorithm for the area-modulating stimulus has been established, a further investigation of test-retest variability would be advisable. As indicated in the study by Artes et al. (2002a), knowledge of the test-retest characteristics are necessary in the determination of early disease from normal, and disease progression from disease stability. These characteristics are also necessary in establishing 95% confidence limits as part of the normative database, and subsequently in the development of software to aid in visual field interpretation, for example in TD probability maps, or progression analysis.

7.4.3.3 Longitudinal study

As indicated in chapter four, the systematic enlargement of the difference in SNR observed between the A and GIII stimuli across the three disease strata, holds promise in the improved identification of disease progression. However, this conclusion cannot be drawn from the current study, as it did not involve longitudinal testing in those with established disease.

Repeated testing of those patients with established glaucoma over several years, with both the area-modulating stimulus, and the standard Goldmann III stimulus, would demonstrate whether the area-modulating stimulus is advantageous in the identification of disease progression from disease stability.

7.4.4   Robustness to optical imperfections

As detailed in chapter six, no statistically significant differences were found between stimulus forms with the addition of positive, spherical lenses to induce optical defocus, however it would be beneficial to expand on this experiment. Further examination with smaller increments of induced blur may demonstrate a difference between the area-modulating stimulus and a standard Goldmann III, with respect to their robustness to optical defocus. Although not a straightforward task, as discussed in chapter six, an investigation into the effects of optical defocus with increasing eccentricity from fixation could also be beneficial in quantifying the area-modulating stimulus.

There were some indications that the area-modulating stimulus was more robust to intraocular straylight than a standard Goldmann III, however quantification of straylight thus far has only been established in the central visual field. A further investigation of the effects of intraocular straylight, with quantification of the peripheral straylight, is therefore recommended, although this would not be a straightforward task.

While these recommendations for further development of this area-modulating stimulus may be ambitious, they would ensure a robust, scientific justification for any introduction of this stimulus into a clinical setting. As detailed in chapter one, the scientific justifications for several parameters currently utilised in SAP, are unclear, or unfounded, and previous attempts at perimetric improvement have been introduced to the clinical setting before a complete investigation had been undertaken. It is therefore crucial in the further development of clinical perimetry that high scientific standards are held, to prevent repetition of past mistakes, and ensure that new methods present a real benefit to patients and clinicians alike.

References

Adams, A. J. et al. 1982. Spectral sensitivity and color discrimination changes in glaucoma and glaucoma-suspect patients. *Investigative Ophthalmology and Visual Science* 23(4), 516-524.

Allen, C. S. et al. 2002. Comparison of the frequency doubling technology screening algorithm and the Humphrey 24-2 SITA-FAST in a large eye screening. *Clinical and Experimental Ophthalmology* 30(1), 8-14.

Almasieh, M. and Levin, L. A. 2017. Neuroprotection in glaucoma: animal models and clinical trials. *Annual Review of Vision Science* 3(1), 91-120.

Anderson, A. J. and Johnson, C. A. 2003. Frequency doubling technology perimetry and optical defocus. *Investigative Ophthalmology and Visual Science* 44(9), 4147-4152.

Anderson, A. J. et al. 2005. Characteristics of the normative database for the Humphrey Matrix Perimeter. *Investigative Ophthalmology and Visual Science* 46(4), 1540-1548.

Anderson, A. J. et al. 2016. Do intense perimetric stimuli saturate the healthy visual system? *Investigative Ophthalmology and Vision Science* 57(14), 6397-6404.

Anderson, D. R. 2003. Collaborative normal tension glaucoma study. *Current Opinion in Ophthalmology* 14(2), 86-90.

Anderson, D. R. and Patella, V. M. 1999. *Automated Static Perimetry*. Second ed. Mosby.

Anderson, R. S. 1996. The selective effect of optical defocus on detection and resolution acuity in peripheral vision. *Current Eye Research* 15(3), 351-353.

Anderson, R. S. 2006. The psychophysics of glaucoma: improving the structure/function relationship. *Progress in Retinal and Eye Research* 25(1), 79-97.

Anderson, R. S. et al. 1992. Psychophysical localization of the human visual streak. *Optometry and Vision Science* 69(3), 171-174.

Anderson, R. S. et al. 2001. Effect of localized defocus on detection thresholds for different sized targets in the fovea and periphery. *Acta Ophthalmologica Scandinavica* 79(1), 60-63.

Anderson, R. S. et al. 2009. The robustness of various forms of perimetry to different levels of induced intraocular stray light. *Investigative Ophthalmology and Visual Science* 50(8), 4022-4028.

Anderson, S. J. and Hess, R. F. 1990. Post-receptoral undersampling in normal human peripheral vision. *Vision Research* 30(10), 1507-1515.

Armaly, M. F. 1969. The visual field defect and ocular pressure level in open angle glaucoma. *Investigative Ophthalmology and Visual Science* 8(1), 105-124.

Artes, P. H. and Chauhan, B. C. 2009. Signal/noise analysis to compare tests for measuring visual field loss and its progression. *Investigative Ophthalmology and Visual Science* 50(10), 4700-4708.

Artes, P. H. et al. 2002a. Properties of perimetric threshold estimates from Full Threshold, SITA Standard, and SITA Fast strategies. *Investigative Ophthalmology and Visual Science* 43(8), 2654-2659.

Artes, P. H. et al. 2002b. Response time as a discriminator between true- and false-positive responses in supra-threshold perimetry. *Investigative Ophthalmology and Visual Science* 43(1), 129-132.

Artes, P. H. et al. 2003a. Multisampling supra-threshold perimetry: a comparison with conventional supra-threshold and full-threshold strategies by computer simulation. *Investigative Ophthalmology and Visual Science* 44(6), 2582-2587.

Artes, P. H. et al. 2003b. Effects of blur and repeated testing on sensitivity estimates with frequency doubling perimetry. *Investigative Ophthalmology and Visual Science* 44(2), 646-652.

Artes, P. H. et al. 2005a. Visual field progression in glaucoma: total versus pattern deviation analyses. *Investigative Ophthalmology and Visual Science* 46(12), 4600-4606.

Artes, P. H. et al. 2005b. Threshold and variability properties of matrix frequency doubling technology and standard automated perimetry in glaucoma. *Investigative Ophthalmology and Visual Science* 46(7), 2451-2457.

Artes, P. H. et al. 2010. Longitudinal and cross-sectional analyses of visual field progression in participants of the ocular hypertension treatment study. *Archives of Ophthalmology* 128(12), 1528-1532.

Artes, P. H. et al. 2014. Visual field progression in glaucoma: what is the specificity of the Guided Progression Analysis? *Ophthalmology* 121(10), 2023-2027.

Asaoka, R. et al. 2012. A novel distribution of visual field test points to improve the correlation between structure-function measurements. *Investigative Ophthalmology and Visual Science* 53(13), 8396-8404.

Asman, P. and Heijl, A. 1992. Glaucoma Hemifield Test. Automated visual field evaluation. *Archives of Ophthalmology* 110(6), 812-819.

Atchison, D. A. 1987. Effect of defocus on visual field measurement. *Ophthalmic and Physiological Optics* 7(3), 259-265.

Atchison, D. A. et al. 1980. Prismatic effects of spherical ophthalmic lenses. *American Journal of Optometry and Physiological Optics* 57(11), 779-790.

Atchison, D. A. et al. 2006. Peripheral refraction along the horizontal and vertical visual fields in myopia. *Vision Research* 46(8), 1450-1458.

Aubert and Foerster. 1857. Untersuchungen über den raumsinn der retina. *Archiv für Ophthalmologie* 3(2), 1-37.

Augsburger, A. and Terry, J. E. 1977. Non-contact and Mackay-Marg tonometry: comparison in patients ages 7 to 85 years. *American Journal of Optometry and Physiological Optics* 54(1), 31-34.

Bach, M. et al. 1997. Raster-scan cathode-ray tubes for vision research-limits of resolution in space, time and intensity, and some solutions. *Spatial Vision* 10(4), 403-414.

Balazsi, A. G. et al. 1984. The effect of age on the nerve fiber population of the human optic nerve. *American Journal of Ophthalmology* 97(6), 760-766.

Barlow, H. B. 1953. Summation and inhibition in the frog's retina. *The Journal of Physiology* 119(1), 69-88.

Barlow, H. B. 1957. Increment thresholds at low intensities considered as signal/noise discriminations. *The Journal of Physiology* 136(3), 469-488.

Barlow, H. B. 1958. Temporal and spatial summation in human vision at different background intensities. *The Journal of Physiology* 141(2), 337-350.

Bates, D. et al. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1-48.

Beck, R. W. et al. 1985. A clinical comparison of visual field testing with a new automated perimeter, the Humphrey Field Analyzer, and the Goldmann perimeter. *Ophthalmology* 92(1), 77-82.

Bellezza, A. J. et al. 2000. The optic nerve head as a biomechanical structure: initial finite element modeling. *Investigative Ophthalmology and Visual Science* 41(10), 2991-3000.

Benedetto, M. D. and Cyrlin, M. N. 1985. The effect of blur upon static perimetric thresholds. In: Heijl, A. and Greve, E.L. eds. *Sixth International Visual Field Symposium*. Dordrecht: Springer Netherlands, 563-567.

Bengtsson, B. and Heijl, A. 1998a. Evaluation of a new perimetric threshold strategy, SITA, in patients with manifest and suspect glaucoma. *Acta Ophthalmologica Scandinavica* 76(3), 268-272.

Bengtsson, B. and Heijl, A. 1998b. SITA Fast, a new rapid perimetric threshold test. Description of methods and evaluation in patients with manifest and suspect glaucoma. *Acta Ophthalmologica Scandinavica* 76(4), 431-437.

Bengtsson, B. and Heijl, A. 2000. False negative responses in glaucoma perimetry: indicators of patient performance or test reliability? *Investigative Ophthalmology and Visual Science* 41(8), 2201-2204.

Bengtsson, B. et al. 1997a. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmologica Scandinavica* 75(4), 368-375.

Bengtsson, B. et al. 1997b. Perimetric probability maps to separate change caused by glaucoma from that caused by cataract. *Acta Ophthalmologica Scandinavica* 75(2), 184-188.

Bengtsson, B. et al. 1998. Evaluation of a new threshold visual field strategy, SITA, in normal subjects. *Acta Ophthalmologica Scandinavica* 76(2), 165-169.

Bergin, C. et al. 2011. The effect of induced intraocular straylight on perimetric tests. *Investigative Ophthalmology and Visual Science* 52(6), 3676-3682.

Berkson, J. 1951. Why I prefer logits to probits. *Biometrics* 7(4), 327-339.

Bloch, A. M. 1885. Experiences sur la vision. *Comptes Rendus de la Societé de Biologie* 37, 493-495.

Blum, F. G. et al. 1959. How important are peripheral fields? *American Medical Association. Archives of Ophthalmology* 61(1), 1-8.

Bode, S. F. et al. 2011. Pattern electroretinogram in glaucoma suspects: new findings from a longitudinal study. *Investigative Ophthalmology and Visual Science* 52(7), 4300-4306.

Boycott, B. B. et al. 1969. Organization of the primate retina: light microscopy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 255, 109-184.

Brainard, D. H. et al. 2002. Display characterization. *Encyclopedia of Imaging Science and Technology.* John Wiley & Sons, Inc.

Brenton, R. S. and Phelps, C. D. 1986. The normal visual field on the Humphrey Field Analyzer. *Ophthalmologica* 193(1-2), 56-74.

Bridgeman, B. 1998. Durations of stimuli displayed on video display terminals:(n-1)/f+ persistence. *Psychological Science* 9(3), 232-233.

Bruder, G. E. and Kietzman, M. L. 1973. Visual temporal integration for threshold, signal detectability, and reaction time measures. *Attention, Perception, & Psychophysics* 13(2), 293-300.

Budenz, D. L. et al. 1993. The effect of simulated cataract on the glaucomatous visual field. *Ophthalmology* 100(4), 511-517.

Burgoyne, C. F. et al. 2005. The optic nerve head as a biomechanical structure: a new paradigm for understanding the role of IOP-related stress and strain in the pathophysiology of glaucomatous optic nerve head damage. *Progress in Retinal and Eye Research* 24(1), 39-73.

Camp, A. S. and Weinreb, R. N. 2017. Will perimetry be performed to monitor glaucoma in 2025? *Ophthalmology* 124(12), S71-S75.

Campbell, F. W. and Green, D. G. 1965. Optical and retinal factors affecting visual resolution. *The Journal of Physiology* 181(3), 576-593.

Caprioli, J. 2001. Should we use short-wavelength automated perimetry to test glaucoma patients? *American Journal of Ophthalmology* 131(6), 792-794.

Casagrande, V. A. 1994. A third parallel visual pathway to primate area V1. *Trends in Neurosciences* 17(7), 305-310.

Casson, R. J. and James, B. 2006. Effect of cataract on frequency doubling perimetry in the screening mode. *Journal of Glaucoma* 15(1), 23-25.

Casson, R. J. et al. 2012. Definition of glaucoma: clinical and experimental concepts. *Clinical and Experimental Ophthalmology* 40(4), 341-349.

Cello, K. E. et al. 2000. Frequency doubling technology perimetry for detection of glaucomatous visual field loss. *American Journal of Ophthalmology* 129(3), 314-322.

Cervino, A. et al. 2008. Performance of the compensation comparison method for retinal straylight measurement: effect of patient's age on repeatability. *The British Journal of Ophthalmology* 92(6), 788-791.

Charman, W. N. and Jennings, J. A. M. 2006. Longitudinal changes in peripheral refraction with age. *Ophthalmic and Physiological Optics* 26(5), 447-455.

Chauhan, B. C. and House, P. H. 1991. Intratest variability in conventional and high-pass resolution perimetry. *Ophthalmology* 98(1), 79-83.

Chauhan, B. C. and Johnson, C. A. 1999. Test-retest variability of frequency doubling perimetry and conventional perimetry in glaucoma patients and normal subjects. *Investigative Ophthalmology and Visual Science* 40(3), 648-656.

Chauhan, B. C. et al. 1993. Characteristics of frequency-of-seeing curves in normal subjects, patients with suspected glaucoma, and patients with glaucoma. *Investigative Ophthalmology and Visual Science* 34(13), 3534-3540.

Chauhan, B. C. et al. 1997. Repeatable diffuse visual field loss in open-angle glaucoma. *Ophthalmology* 104(3), 532-538.

Chong, G. T. and Lee, R. K. 2012. Glaucoma versus red disease: imaging and glaucoma diagnosis. *Current Opinion in Ophthalmology* 23(2), 79-88.

Choplin, N. T. et al. 1990. The effect of stimulus size on the measured threshold values in automated perimetry. *Ophthalmology* 97(3), 371-374.

Chylack, L. T. et al. 1993. The lens opacities classification system III. *Archives of Ophthalmology* 111(6), 831-836.

Colton, T. and Ederer, F. 1980. The distribution of intraocular pressures in the general population. *Survey of Ophthalmology* 25(3), 123-129.

Comerford, J. P. et al. 2002. Intensity discrimination. In: Norton, T.T. et al. eds. *The Psychophysical Measurement of Visual Function*. Woburn: Butterworth-Heinemann.

Cooper, E. A. et al. 2013. Assessment of OLED displays for vision research. *Journal of Vision* 13(12):16, 1-13.

Corliss, D. A. and Norton, T. T. 2002. Principles of Psychophysical Measurement. In: Norton, T.T. et al. eds. *The Psychophysical Measurement of Visual Function*. Woburn: Butterworth-Heinemann.

Costagliola, C. et al. 1990. Intraocular pressure in a healthy population: a survey of 751 subjects. *Optometry and Vision Science* 67(3), 204-206.

Cowan, W. 1995. Displays for vision research. In: Bass, M. et al. eds. *Handbook of Optics, Volume I: Fundamentals, techniques, & design*. New York: McGraw-Hill.

Crawford, M. L. J. et al. 2000. Glaucoma in Primates: Cytochrome Oxidase Reactivity in Parvo- and Magnocellular Pathways. *Investigative Ophthalmology and Visual Science* 41(7), 1791-1802.

Crish, S. D. et al. 2010. Distal axonopathy with structural persistence in glaucomatous neurodegeneration. *Proceedings of the National Academy of Sciences of the United States of America* 107(11), 5196-5201.

Crosswell, H. H. et al. 1991. The effect of background intensity on the components of fluctuation as determined by threshold-related automated perimetry. *Graefe's Archive for Clinical and Experimental Ophthalmology* 229(2), 119-122.

Curcio, C. A. and Drucker, D. N. 1993. Retinal ganglion cells in Alzheimer's disease and aging. *Annals of Neurology* 33(3), 248-257.

Curcio, C. A. et al. 1990. Human photoreceptor topography. *The Journal of Comparative Neurology* 292(4), 497-523.

Curcio, C. A. et al. 1993. Aging of the human photoreceptor mosaic: evidence for selective vulnerability of rods in central retina. *Investigative Ophthalmology and Visual Science* 34(12), 3278-3296.

Dalimier, E. and Dainty, C. 2010. Role of ocular aberrations in photopic spatial summation in the fovea. *Optics Letters* 35(4), 589-591.

Dandona, L. et al. 1991. Selective effects of experimental glaucoma on axonal transport by retinal ganglion cells to the dorsal lateral geniculate nucleus. *Investigative Ophthalmology and Visual Science* 32(5), 1593-1599.

Dannheim, F. and Drance, S. M. 1974. Psychovisual disturbances in glaucoma. A study of temporal and spatial summation. *Archives of Ophthalmology* 91(6), 463-468.

Davanger, M. and Holter, Ö. 1965. The statistical distribution of intraocular pressure in the population. *Acta Ophthalmologica* 43(2), 314-322.

Davidson-Hall, T. et al. 2017. Organic light emitting device materials for displays. In: Kitai, A. ed. *Materials for Solid State Lighting and Displays*. John Wiley & Sons, 183-230.

Davila, K. D. and Geisler, W. S. 1991. Relative contributions of pre-neural and neural factors to areal summation in the fovea. *Vision Research* 31(7/8), 1369-1380.

De Waard, P. W. et al. 1992. Intraocular light scattering in age-related cataracts. *Investigative Ophthalmology and Visual Science* 33(3), 618-625.

DeLeón-Ortega, J. E. et al. 2006. Discrimination between glaucomatous and nonglaucomatous eyes using quantitative imaging devices and subjective optic nerve head assessment. *Investigative Ophthalmology and Visual Science* 47(8), 3374-3380.

Demirel, S. and Johnson, C. A. 2000. Isolation of short-wavelength sensitive mechanisms in normal and glaucomatous visual field regions. *Journal of Glaucoma* 9(1), 63-73.

Drance, S. M. 1969. The early field defects in glaucoma. *Investigative Ophthalmology and Visual Science* 8(1), 84-91.

Drance, S. M. 1991. Diffuse visual field loss in open-angle glaucoma. *Ophthalmology* 98(10), 1533-1538.

Drews-Bankiewicz, M. A. et al. 1992. Contrast sensitivity in patients with nuclear cataracts. *Archives of Ophthalmology* 110(7), 953-959.

Dubois-Poulsen, A. et al. 1952. *Le champ visuel*. Masson, Paris.

EGS. 2014. *Terminology and guidelines for glaucoma*. 4th ed. Savona, Italy.

Ehlers, H. 1981. Bjerrum's clinic. *Survey of Ophthalmology* 26(2), 101-102.

Elliott, D. B. et al. 1989. Contrast sensitivity and glare sensitivity changes with three types of cataract morphology: are these techniques necessary in a clinical evaluation of cataract? *Ophthalmic and Physiological Optics* 9(1), 25-30.

Elliott, D. B. 1993. Evaluating visual function in cataract. *Optometry and Vision Science* 70(11), 896-902.

Elliot, R. H. 1920. A lecture on the diagnosis of glaucoma: given at the North-East London Post-Graduate College, Tottenham, January 20th, 1920. *British Medical Journal* 1(3087), 279-282.

Elze, T. 2010a. Achieving precise display timing in visual neuroscience experiments. *Journal of Neuroscience Methods* 191(2), 171-179.

Elze, T. 2010b. Misspecifications of stimulus presentation durations in experimental psychology: a systematic review of the psychophysics literature. *PLoS One* 5(9), e12792.

Elze, T. et al. 2013. An evaluation of organic light emitting diode monitors for medical applications: great timing, but luminance artifacts. *Medical Physics* 40(9), 092701.

Ennis, F. A. and Anderson, R. S. 2000. Aliasing in peripheral vision for flickering gratings under different levels of illumination. *Current Eye Research* 20(5), 413-419.

Fankhauser, F. 1979. Problems related to the design of automatic perimeters. *Documenta Ophthalmologica* 47(1), 89-138.

Fankhauser, F. and Enoch, J. M. 1962. The effects of blur upon perimetric thresholds. A method for determining a quantitative estimate of retinal contour. *Archives of Ophthalmology* 68(2), 240-251.

Fankhauser, F. and Schmidt, T. 1958. Die untersuchung der räumlichen summation mit stehender und bewegter reizmarke nach der methode der quantitativen lichtsinnperimetrie. *Ophthalmologica* 135(5-6), 660-666.

Fankhauser, F. and Schmidt, T. 1960. Die optimalen bedingungen für die untersuchung der räumlichen summation mit stehender reizmarke nach der methode der quantitativen lichtsinnperimetrie. *Ophthalmologica* 139(5), 409-423.

Fankhauser, F. et al. 1977. Some aspects of the automation of perimetry. *Survey of Ophthalmology* 22(2), 131-141.

Farrell, J. E. et al. 2017. Characterization of visual stimuli using the standard display model. In: Artal, P. ed. *Handbook of Visual Optics, Volume One: Fundamentals and Eye Optics*. CRC Press, Taylor & Francis Group, 93-102.

Fechtner, R. D. and Weinreb, R. N. 1994. Mechanisms of optic nerve damage in primary open angle glaucoma. *Survey of Ophthalmology* 39(1), 23-42.

Felius, J. et al. 1995. Functional characteristics of blue-on-yellow perimetric thresholds in glaucoma. *Investigative Ophthalmology and Visual Science* 36(8), 1665-1674.

Felius, J. et al. 1997. Spatial summation for selected ganglion cell mosaics in patients with glaucoma. In: Wall, M. and Heijl, A. eds. *Perimetry Update 1996/1997 Proceedings of the XIIth International Perimetric Society Meeting.* Würzburg, Germany. Amsterdam/New York: Kugler, 213-221.

Fellman, R. L. et al. 1988. Clinical importance of spatial summation in glaucoma. In: Heijl, A. ed. *VIII International Perimetric Society Meeting.* Vancouver, Canada. Amsterdam: Kulger & Ghedini Publications, 313-324.

Ferree, C. E. and Rand, G. 1920a. Factors which influence the color sensitivity of the peripheral retina. *Transactions of the American Ophthalmological Society* 18, 244-271.

Ferree, C. E. and Rand, G. 1920b. The campperimeter-an illuminated perimeter with campimeter features. *Transactions of the American Ophthalmological Society* 18, 164-172.

Ferree, C. E. and Rand, G. G. 1933. Interpretation of refractive conditions in the peripheral field of vision: A further study. *Archives of Ophthalmology* 9(6), 925-938.

Ferree, C. E. et al. 1931. Refraction for the peripheral field of vision. *Archives of Ophthalmology* 5(5), 717-731.

Fischer, B. 1973. Overlap of receptive field centres and representation of the visual fields in the cat's optic tract. *Vision Research* 13, 2113-2120.

Flammer, J. et al. 1984. Differential light threshold: Short- and long-term fluctuation in patients with glaucoma, normal controls, and patients with suspected glaucoma. *Archives of Ophthalmology* 102(5), 704-706.

Foerster, R. 1869. Vgl. über gesichtsfeldmessung. *Klinische Monatsblätter für Augenheilkunde* 7, 411-415.

Fonn, D. et al. 1999. Sympathetic swelling response of the control eye to soft lenses in the other eye. *Investigative Ophthalmology and Visual Science* 40(13), 3116-3121.

Franssen, L. et al. 2006. Compensation comparison method for assessment of retinal straylight. *Investigative Ophthalmology and Visual Science* 47(2), 768-776.

Fredette, M.-J. et al. 2015. Comparison of matrix with Humphrey Field Analyzer II with SITA. *Optometry and Vision Science* 92(5), 527-536.

Friedmann, A. I. 1966. Serial analysis of changes in visual field defects, employing a new instrument, to determine the activity of diseases involving the visual pathways. *Ophthalmologica* 152(1), 1-12.

Frisén, L. 1987. High-pass resolution targets in peripheral vision. *Ophthalmology* 94(9), 1104-1108.

Fung, S. S. et al. 2013. Are practical recommendations practiced? A national multi-centre cross-sectional study on frequency of visual field testing in glaucoma. *British Journal of Ophthalmology* 97(7), 843-847.

Gafner, F. and Goldmann, H. 1955. Experimentelle untersuchungen über den zusammenhang von augendrucksteigerung und gesichtsfeldschädigung. *Ophthalmologica* 130(6), 357-377.

Gardiner, S. K. and Crabb, D. P. 2002. Frequency of testing for detecting visual field progression. *British Journal of Ophthalmology* 86(5), p. 560.

Gardiner, S. K. et al. 2008. Is there evidence for continued learning over multiple years in perimetry? *Optometry and Vision Science* 85(11), 1043-1048.

Gardiner, S. K. et al. 2014. Assessment of the reliability of standard automated perimetry in regions of glaucomatous damage. *Ophthalmology* 121(7), 1359-1369.

Gardiner, S. K. et al. 2015. The effect of stimulus size on the reliable stimulus range of perimetry. *Translational Vision Science and Technology* 4(2), 10.

Garway-Heath, D. F. et al. 2000a. Mapping the visual field to the optic disc in normal tension glaucoma eyes. *Ophthalmology* 107(10), 1809-1815.

Garway-Heath, D. F. et al. 2000b. Scaling the hill of vision: the physiological relationship between light sensitivity and ganglion cell numbers. *Investigative Ophthalmology and Visual Science* 41(7), 1774-1782.

Garway-Heath, D. F. et al. 2002. Relationship between electrophysiological, psychophysical, and anatomical measurements in glaucoma. *Investigative Ophthalmology and Vision Science* 43(7), 2213-2220.

Geffroy, B. et al. 2017. Organic light-emitting diode (OLED) technology: materials, devices and display technologies. *Polymer International* 55(6), 572-582.

Gerente, V. M. et al. 2015. Evaluation of glaucomatous damage via functional magnetic resonance imaging, and correlations thereof with anatomical and psychophysical ocular findings. *PLoS One* 10(5), E0126362.

Ghodrati, M. et al. 2015. The (un)suitability of modern liquid crystal displays (LCDs) for vision research. *Frontiers in Psychology* 6, 1-11.

Gilbert, C. D. and Wiesel, T. N. 1992. Receptive field dynamics in adult primary visual cortex. *Nature* 356, 150-152.

Gilchrist, J. M. et al. 2005. Comparing and unifying slope estimates across psychometric function models. *Perception and Psychophysics* 67(7), 1289-1303.

Gilpin, L. B. et al. 1990. Threshold variability using different Goldmann stimulus sizes. *Acta Ophthalmologica (Copenhagen)* 68(6), 674-676.

Glen, F. C. et al. 2014. A qualitative investigation into patients' views on visual field testing for glaucoma monitoring. *British Medical Journal Open* 4(1), E003996.

Glezer, V. D. 1965. The receptive fields of the retina. *Vision Research* 5(10-11), 497-525.

Gloor, B. P. 2009. Franz Fankhauser: the father of the automated perimeter. *Survey of Ophthalmology* 54(3), 417-425.

Goldmann, H. 1945a. Grundlagen exakter perimetrie. *Ophthalmologica* 109(2-3), 57-70.

Goldmann, H. 1945b. Ein selbstregistrierendes projektionskugelperimeter. *Ophthalmologica* 109(2-3), 71-79.

Goldmann, H. 1946. Demonstration unseres neuen projektionskugelperimeters samt theoretischen und klinischen bemerkungen über perimetrie. *Ophthalmologica* 111(2-3), 187-192.

Goldmann, H. 1999. Fundamentals of exact perimetry. *Optometry and Vision Science* 76(8), 599-604.

Golemez, H. et al. 2016. Is multifocal electroretinography an early predictor of glaucoma? *Documenta Ophthalmologica* 132(1), 27-37.

Graham, C. H. and Bartlett, N. R. 1939. The relation of size of stimulus and intensity in the human eye: II. Intensity thresholds for red and violet light. *Journal of Experimental Psychology* 24(6), 574-587.

Graham, C. H. et al. 1939. The relation of size of stimulus and intensity in the human eye: I. Intensity thresholds for white light. *Journal of Experimental Psychology* 24(6), 555-573.

Greve, E. L. 1973. Single and multiple stimulus static perimetry in glaucoma; the two phases of perimetry. *Documenta Ophthalmologica* 36(1), 1-346.

Grobbel, J. et al. 2016. Normal values for the full visual field, corrected for age- and reaction time, using semiautomated kinetic testing on the Octopus 900 perimeter. *Translational Vision Science and Technology* 5(2):5, 1-13.

Guber, I. et al. 2011. Reproducibility of straylight measurement by C-Quant for assessment of retinal straylight using the compensation comparison method. *Graefe's Archive for Clinical and Experimental Ophthalmology* 249(9), 1367-1371.

Haag-Streit. 2014. Octopus 900, Flexibility and reliability [Brochure]. *Haag-Streit AG, Ophthalmological Instruments 3098 Köniz/Switzerland*.

Haag-Streit AG. n.d. Original Goldmann Perimeter 940 [Brochure]. *Haag-Streit AG, Ophthalmological Instruments 3098 Köniz/Switzerland*.

Haas, A. et al. 1986. Influence of age on the visual fields of normal subjects. *American Journal of Ophthalmology* 101(2), 199-203.

Hallett, P. E. 1963. Spatial summation. *Vision Research* 3(1), 9-24.

Harrington, D. O. 1964. The Bjerrum scotoma. *Transactions of the American Ophthalmological Society* 62, 324-348.

Hartline, H. K. 1938. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology--Legacy Content* 121(2), 400-415.

Hartline, H. K. 1940. The receptive fields of optic nerve fibers. *American Journal of Physiology* 130(4), 690-699.

Harvey, L. O. 1986. Efficient estimation of sensory thresholds. *Behavior Research Methods* 18(6), 623-632.

Harwerth, R. S. et al. 1999. Ganglion cell losses underlying visual field defects from experimental glaucoma. *Investigative Ophthalmology and Visual Science* 40(10), 2242-2250.

Hayashi, K. et al. 2001. Influence of cataract surgery on automated perimetry in patients with glaucoma. *American Journal of Ophthalmology* 132(1), 41-46.

Haymes, S. A. et al. 2005. Glaucomatous visual field progression with frequency doubling technology and standard automated perimetry in a longitudinal prospective study. *Investigative Ophthalmology and Visual Science* 46(2), 547-554.

Hayreh, S. S. et al. 1970. Vasogenic origin of visual field defects and optic nerve changes in glaucoma. *The British Journal of Ophthalmology* 54(7), 461-472.

Heijl, A. 1976. Automatic perimetry in glaucoma visual field screening. *Graefe's Archive for Clinical and Experimental Ophthalmology* 200(1), 21-37.

Heijl, A. 1984. The Humphrey Field Analyzer, construction and concepts. In: Henkes, H.E. ed. *Sixth International Visual Fields Symposium, Documenta Ophthalmologica.* Santa Margherita Ligure. Dr W. Junk Publishers, 77-84.

Heijl, A. and Bengtsson, B. 1996. The effect of perimetric experience in patients with glaucoma. *Archives of Ophthalmology* 114(1), 19-22.

Heijl, A. and Krakau, C. E. 1975. An automatic static perimeter, design and pilot study. *Acta Ophthalmologica* 53(3), 293-310.

Heijl, A. et al. 1987. Normal variability of static perimetric threshold values across the central visual field. *Archives of Ophthalmology* 105(11), 1544-1549.

Heijl, A. et al. 1989a. Test-retest variability in glaucomatous visual fields. *American Journal of Ophthalmology* 108(2), 130-135.

Heijl, A. et al. 1989b. The effect of perimetric experience in normal subjects. *Archives of Ophthalmology* 107(1), 81-86.

Heijl, A. et al. 2009. Natural history of open-angle glaucoma. *Ophthalmology* 116(12), 2271-2276.

Heijl, A. et al. 2012. The field analyzer primer: Effective perimetry. Carl Zeiss Meditec, Inc.

Hennelly, M. L. et al. 1998. The effect of age on the light scattering characteristics of the eye. *Ophthalmic and Physiological Optics* 18(2), 197-203.

Henson, D. B. and Morris, E. J. 1993. Effect of uncorrected refractive errors upon central visual field testing. *Ophthalmic and Physiological Optics* 13(4), 339-343.

Henson, D. B. et al. 1984. Evaluation of the Friedmann Visual Field Analyser Mark II. Part 1. Results from a normal population. *The British Journal of Ophthalmology* 68(7), 458-462.

Henson, D. B. et al. 1999. Diffuse loss of sensitivity in early glaucoma. *Investigative Ophthalmology and Visual Science* 40(13), 3147-3151.

Henson, D. B. et al. 2000. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Investigative Ophthalmology and Visual Science* 41(2), 417-421.

Heron, G. et al. 1988. Central visual fields for short wavelength sensitive pathways in glaucoma and ocular hypertension. *Investigative Ophthalmology and Visual Science* 29(1), 64-72.

Herse, P. R. 1992. Factors influencing normal perimetric thresholds obtained using the Humphrey Field Analyzer. *Investigative Ophthalmology and Visual Science* 33(3), 611-617.

Herse, P. et al. 1998. Factors influencing short wavelength automated perimetry in normal subjects. *Clinical and Experimental Optometry* 81(2), 77-80.

Hess, R. and Woo, G. 1978. Vision through cataracts. *Investigative Ophthalmology & Visual Science* 17(5), 428-435.

Heuer, D. K. et al. 1987. The influence of refraction accuracy on automated perimetric threshold measurements. *Ophthalmology* 94(12), 1550-1553.

Heuer, D. K. et al. 1988. The influence of simulated light scattering on automated perimetric threshold measurements. *Archives of Ophthalmology* 106(9), 1247-1251.

Hill, N. J. 2001. *Testing hypotheses about psychometric functions.* University of Oxford.

Hirasawa, K. et al. 2016. Comparison of size modulation and conventional standard automated perimetry with the 24-2 test protocol in glaucoma patients. *Scientific Reports* 6: 25563.

Hodapp, E. et al. 1993. Clinical decisions in glaucoma. *Mosby, St. Louis*, 52-61.

Horner, D. G. et al. 2013. Blur-resistant perimetric stimuli. *Optometry and Vision Science* 90(5), 466-474.

Hoyt, W. F. and Newman, N. M. 1972. The earliest observable defect in glaucoma? *The Lancet* 1, 692-693.

Hulke, J. W. 1858. On some points in the pathology and morbid anatomy of glaucoma. *Medico-Chirurgical Transactions* 41, 111-118.

Ijspeert, J. K. et al. 1990. The intraocular straylight function in 129 healthy volunteers; Dependence on angle, age and pigmentation. *Vision Research* 30(5), 699-707.

Ikeda, H. and Wright, M. J. 1972. Receptive field organization of 'sustained' and 'transient' retinal ganglion cells which subserve different functional roles. *Journal of Physiology* 227(3), 769-800.

Ito, H. et al. 2013. Evaluation of an organic light-emitting diode display for precise visual stimulation. *Journal of Vision* 13(7): 6, 1-21.

Jacobs, G. H. 1969. Receptive fields in visual systems. *Brain Research* 14(3), 553-573.

Jafarzadehpour, E. et al. 2013. Pattern electroretinography in glaucoma suspects and early primary open angle glaucoma. *Journal of Ophthalmic and Vision Research* 8(3), 199-206.

Je, S. et al. 2018. Spatial summation across the visual field in strabismic and anisometropic amblyopia. *Scientific Reports* 8(1), 3858.

Jeaffreson, C. S. 1873. Notes on a new perimeter. *British Medical Journal* 2(678), 752-755.

Johnson, C. A. 1994. Selective versus nonselective losses in glaucoma. *Journal of glaucoma* 3, S32-44.

Johnson, C. A. 1996. Diagnostic value of short-wavelength automated perimetry. *Current Opinion in Ophthalmology* 7(2), 54-58.

Johnson, C. A. and Samuels, S. J. 1997. Screening for glaucomatous visual field loss with frequency doubling perimetry. *Investigative Ophthalmology and Visual Science* 38(2), 413-425.

Johnson, C. A. et al. 1993a. Blue-on-yellow perimetry can predict the development of glaucomatous visual field loss. *Archives of Ophthalmology* 111(5), 645-650.

Johnson, C. A. et al. 1993b. Progression of early glaucomatous visual field loss as detected by blue-on-yellow and standard white-on-white automated perimetry. *Archives of Ophthalmology* 111(5), 651-656.

Johnson, C. A. et al. 1995. Short-wavelength automated perimetry in low-, medium-, and high-risk ocular hypertensive eyes. Initial baseline results. *Archives of Ophthalmology* 113(1), 70-76.

Johnson, C. A. et al. 2011. A history of perimetry and visual field testing. *Optometry and Vision Science* 88(1), E8-E15.

Kalloniatis, M. and Khuu, S. K. 2016. Equating spatial summation in visual field testing reveals greater loss in optic nerve disease. *Ophthalmic and Physiological Optics* 36(4), 439-452.

Kalmus, H. et al. 1974. Impairment of colour vision in patients with ocular hypertension and glaucoma. With special reference to the "D and H color-rule". *The British Journal of Ophthalmology* 58(11), 922-926.

Kass, M. A. et al. 2002. The ocular hypertension treatment study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Archives of Ophthalmology* 120(6), 701-713.

Katz, J. and Sommer, A. 1986. Asymmetry and variation in the normal hill of vision. *Archives of Ophthalmology* 104(1), 65-68.

Katz, J. and Sommer, A. 1987. A longitudinal study of the age-adjusted variability of automated visual fields. *Archives of Ophthalmology* 105(8), 1083-1086.

Kelly, D. 1966. Frequency doubling in visual responses. *Journal of the Optical Society of America* 56(11), 1628-1633.

Kerrigan-Baumrind, L. A. et al. 2000. Number of ganglion cells in glaucoma eyes compared with threshold visual field tests in the same persons. *Investigative Ophthalmology and Visual Science* 41(3), 741-748.

Khuu, S. K. and Kalloniatis, M. 2015a. Spatial summation across the central visual field: Implications for visual field testing. *Journal of Vision* 15(1), 1-15.

Khuu, S. K. and Kalloniatis, M. 2015b. Standard automated perimetry: determining spatial summation and its effect on contrast sensitivity across the visual field. *Investigative Ophthalmology and Visual Science* 56(6), 3565-3576.

Kim, Y. Y. et al. 2001. Effect of cataract extraction on blue-on-yellow visual field. *American Journal of Ophthalmology* 132(2), 217-220.

King-Smith, P. E. et al. 1994. Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Research* 34(7), 885-912.

Kingdom, F. and Prins, N. 2009. *Psychophysics: A Practical Introduction*. Academic Press.

Klein, J. et al. 2013. Photometric and colorimetric measurements of CRT and TFT monitors for vision research. *Journal of Modern Optics* 60(14), 1159-1166.

Knoblauch, K. 2014. psyphy*: Functions for analyzing psychophysical data in R. R package version 0.1-9*, http://CRAN.R-project.org/package=psyphy.

Kolko, M. 2015. Suppl 1: M5: Present and new treatment strategies in the management of glaucoma. *The Open Ophthalmology Journal* 9, 89-100.

Kook, M. S. et al. 2004. Effect of cataract extraction on frequency doubling technology perimetry. *American Journal of Ophthalmology* 138(1), 85-90.

Krantz, J. 2000. Tell me, what did you see? The stimulus on computers. *Behavior Research Methods, Instruments, & Computers* 32(2), 221-229.

Kuffler, S. W. 1953. Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology* 16(1), 37-68.

Kwon, Y. H. et al. 1998. Test-retest variability of blue-on-yellow perimetry is greater than white-on-white perimetry in normal subjects. *American Journal of Ophthalmology* 126(1), 29-36.

Lam, B. L. et al. 1991. Effect of cataract on automated perimetry. *Ophthalmology* 98(7), 1066-1070.

Larrosa, J. M. et al. 2000. Short-wavelength automated perimetry and neuroretinal rim area. *European Journal of Ophthalmology* 10(2), 116-120.

Latham, K. et al. 1993. Magnification perimetry. *Investigative Ophthalmology & Visual Science* 34(5), 1691-1701.

Latham, K. et al. 1994. Spatial summation of the differential light threshold as a function of visual field location and age. *Ophthalmic & Physiological Optics* 14(1), pp. 71-78.

Lee, B. B. 1996. Receptive field structure in the primate retina. *Vision Research* 36(5), 631-644.

Lelkens, A. M. and Zuidema, P. 1983. Increment thresholds with various low background intensities at different locations in the peripheral retina. *The Journal of the Optical Society of America* 73(10), 1372-1378.

Leydhecker, W. 1959. Zur verbreitung des glaucoma simplex in der scheinbar gesunden, augenärztlich nicht behandelten bevölkerung. *Documenta Ophthalmologica* 13(1), 359-380.

Leydhecker, W. et al. 1958. Der intraokulare druck gesunder menschlicher augen. *Klinische Monatsblatter Fur Augenheilkunde* 133(5), 662-670.

Lie, I. 1981. Visual detection and resolution as a function of adaptation and glare. *Vision Research* 21(12), 1793-1797.

Liu, S. et al. 2011. Comparison of standard automated perimetry, frequency doubling technology perimetry, and short-wavelength automated perimetry for detection of glaucoma. *Investigative Ophthalmology and Visual Science* 52(10), 7325-7331.

Logan, N. and Anderson, D. R. 1983. Detecting early glaucomatous visual field changes with a blue stimulus. *American Journal of Ophthalmology* 95(4), 432-434.

Lynn, J. R. 1969. Examination of the visual field in glaucoma. *Investigative Ophthalmology and Visual Science* 8(1), 76-84.

Lynn, J. R. et al. eds. 1991. Evaluation of automated kinetic perimetry (AKP) with the Humphrey Field Analyzer. *Perimetry update 1990/1991.* Malmö, Sweden. Kugler and Ghedini Publications, 433-454.

Löhle, F. 1929. Über die abhängigkeit des reizschwellenwertes vom sehwinkel. *Zeitschrift für Physik* 54(1-2), 137-151.

Maddess, T. L. and Henry, G. 1992. Performance of nonlinear visual units in ocular hypertension and glaucoma. *Clinical Vision Sciences* 7(5), 371-383.

Malik, R. et al. 2006. Development and evaluation of a linear staircase strategy for the measurement of perimetric sensitivity. *Vision Research* 46(18), 2956-2967.

Malik, R. et al. 2012. 'Structure-function relationship' in glaucoma: past thinking and current concepts. *Clinical and Experimental Ophthalmology* 40(4), 369-380.

Malik, R. et al. 2013. A survey of attitudes of glaucoma subspecialists in England and Wales to visual field test intervals in relation to NICE guidelines. *British Medical Journal Open* 3(5), E002067.

Martin, P. R. et al. 1997. Evidence that blue-on cells are part of the third geniculocortical pathway in primates. *The European Journal of Neuroscience* 9(7), 1536-1541.

Martinez-Conde, S. et al. 2004. The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience* 5(3), 229-240.

May, K. A. and Solomon, J. A. 2013. Four theorems on the psychometric function. *PLoS One* 8(10), E74815.

McKee, S. P. et al. 1985. Statistical properties of forced-choice psychometric functions: Implications of probit analysis. *Attention, Perception, and Psychophysics* 37(4), 286-298.

Meacock, W. R. et al. 2003. The effect of posterior capsule opacification on visual function. *Investigative Ophthalmology and Visual Science* 44(11), 4665-4669.

Medeiros, F. A. et al. 2015. Longitudinal changes in quality of life and rates of progressive visual field loss in glaucoma patients. *Ophthalmology* 122(2), 293-301.

Merigan, W. H. and Maunsell, J. H. 1993. How parallel are the primate visual pathways? *Annual Review of Neuroscience* 16(1), 369-402.

Metha, A. et al. 1993. Calibration of a color monitor for visual psychophysics. *Behavior Research Methods, Instruments, & Computers* 25(3), 371-383.

Millodot, M. 1981. Effect of ametropia on peripheral refraction. *American Journal of Optometry and Physiological Optics* 58(9), 691-695.

Mollon, J. D. 1998. Specifying, generating, and measuring colours. In: Carpenter, R.H.S. and Robson, J.G. eds. *Vision Research: A Practical Guide to Laboratory Methods.* New York: Oxford University Press.

Morgan, J. E. et al. 2000. Retinal ganglion cell death in experimental glaucoma. *British Journal of Ophthalmology* 84(3), 303-310.

Morgan, J. E. and Tribble, J. R. 2015. Microbead models in glaucoma. *Experimental Eye Research* 141, 9-14.

Morrison, J. C. et al. 2008. Rat models for glaucoma research. *Prog Brain Res* 173, 285-301.

Moss, I. D. et al. 1995. The influence of age-related cataract on blue-on-yellow perimetry. *Investigative Ophthalmology and Visual Science* 36(5), 764-773.

Mulholland, P. 2014. *Temporal summation with age and in glaucoma.* University of Ulster.

Mulholland, P. J. et al. 2015a. Estimating the critical duration for temporal summation of standard achromatic perimetric stimuli. *Investigative Ophthalmology and Visual Science* 56(1), 431-437.

Mulholland, P. J. et al. 2015b. Spatiotemporal summation of perimetric stimuli in early glaucoma. *Investigative Ophthalmology and Visual Science* 56(11), 6473-6482.

Mulholland, P. J. et al. 2015c. The effect of age on the temporal summation of achromatic perimetric stimuli. *Investigative Ophthalmology and Visual Science* 56(11), 6467-6472.

Mulholland, P. J. et al. 2015d. Effect of varying CRT refresh rate on the measurement of temporal summation. *Ophthalmic and Physiological Optics* 35(5), 582-590.

Mönter , V. M. et al. 2017. Reclaiming the periphery: automated kinetic perimetry for measuring peripheral visual fields in patients with glaucoma. *Investigative Ophthalmology and Visual Science* 58(2), 868-875.

Neil Charman, W. and Radhakrishnan, H. 2010. Peripheral refraction and the development of refractive error: a review. *Ophthalmic and Physiological Optics* 30(4), 321-338.

NICE. 2009. Glaucoma: Diagnosis and management of chronic open angle glaucoma and ocular hypertension *National Institute for Health and Clinical Excellence*.

NICE. 2017. Glaucoma: diagnosis and management. *National Institute for Health and Care Excellence*.

Nouri-Mahdavi, K. et al. 2011. Detection of visual field progression in glaucoma with standard achromatic perimetry: a review and practical implications. *Graefe's Archive for Clinical and Experimental Ophthalmology* 249(11), 1593-1616.

Nowomiejska, K. et al. 2015. Semi-automated kinetic perimetry provides additional information to static automated perimetry in the assessment of the remaining visual field in end-stage glaucoma. *Ophthalmic and Physiological Optics* 35(2), 147-154.

Orta, A. O. et al. 2015. The correlation between glaucomatous visual field loss and vision-related quality of life. *The Journal of Glaucoma* 24(5), E121-E127.

Owsley, C. et al. 1983. Contrast sensitivity throughout adulthood. *Vision Research* 23(7), 689-699.

Pan, F. and Swanson, W. H. 2006. A cortical pooling model of spatial summation for perimetric stimuli. *Journal of Vision* 6(11), 1159-1171.

Pardhan, S. and Gilchrist, J. 1991. The importance of measuring binocular contrast sensitivity in unilateral cataract. *Eye (London)* 5(1), 31-35.

Pennebaker, G. E. et al. 1992. The effect of stimulus duration upon the components of fluctuation in static automated perimetry. *Eye (London)* 6(4), 353-355.

Perkins, E. S. 1973. The Bedford glaucoma survey. I. Long-term follow-up of borderline cases. *The British Journal of Ophthalmology* 57(3), 179-185.

Phu, J. et al. 2017a. The value of visual field testing in the era of advanced imaging: clinical and psychophysical perspectives. *Clinical and Experimental Optometry* 100(4), 313-332.

Phu, J. et al. 2017b. A comparison of Goldmann III, V and spatially equated test stimuli in visual field testing: the importance of complete and partial spatial summation. *Ophthalmic and Physiological Optics* 37(2), 160-176.

Pickard, R. 1931. Glaucoma and low tension disc cup enlargements *The British Journal of Ophthalmology* 15(6), 323-333.

Pillunat, K. R. et al. 2015. Nocturnal blood pressure in primary open-angle glaucoma. *Acta Ophthalmologica* 93(8), e621-626.

Piper, H. 1903. Über die abhängigkeit des reizwertes leuchtender objekte von ihrer flächen-bezw. winkelgröße. (Fortsetzung der untersuchungen über dunkeladaptation des sehorgans). *Zeitschrift für Psychologie der Sinnesorgane* 32, 98-112.

Piéron, H. 1920. De la valeur de l'énergie liminaire en fonction de la surface retinienne excitée pour la vision fovéale, et de l'influence réciproque de la durée et de la surface d'excitation sur la sommation spatiale ou temporelle pour la vision fovéale et périphérique.(cônes et bâtonnets). *Comptes Rendus Des Seances De La Societe De Biologie Et De Ses Filliales* 83, 1072-1076.

Piéron, H. 1929. V. De la sommation spatiale des impressions lumineuses au niveau de la fovea. *L'Année Psychologique* 30(1), 87-105.

Poinoosawmy, D. et al. 1980. Colour vision in patients with chronic simple glaucoma and ocular hypertension. *The British Journal of Ophthalmology* 64(11), 852-857.

Polo, V. et al. 1998. Short-wavelength automated perimetry and retinal nerve fiber layer evaluation in suspected cases of glaucoma. *Archives of Ophthalmology* 116(10), 1295-1298.

Portney, G. L. and Krohn, M. A. 1978. Automated perimetry: background, instruments and methods. *Survey of Ophthalmology* 22(4), 271-278.

Preiser, D. et al. 2013. Photopic negative response versus pattern electroretinogram in early glaucoma. *Investigative Ophthalmology and Visual Science* 54(2), 1182-1191.

Prins, N. and Kingdom, F. A. A. 2009. Palamedes: Matlab routines for analyzing psychophysical data.

Quigley, H. A. 1985. Early detection of glaucomatous damage: II. Changes in the appearance of the optic disk. *Survey of Ophthalmology* 30(2), 117-126.

Quigley, H. A. et al. 1982. Optic nerve damage in human glaucoma. III. Quantitative correlation of nerve fiber loss and visual field defect in glaucoma, ischemic neuropathy, papilledema, and toxic neuropathy. *Archives of Ophthalmology* 100(1), 135-146.

Quigley, H. A. et al. 1987. Chronic glaucoma selectively damages large optic nerve fibers. *Investigative Ophthalmology and Visual Science* 28(6), 913-920.

Quigley, H. A. et al. 1988. Chronic human glaucoma causing selectively greater loss of large optic nerve fibers. *Ophthalmology* 95(3), 357-363.

Quigley, H. A. et al. 1989. Retinal ganglion cell atrophy correlated with automated perimetry in human eyes with glaucoma. *American Journal of Ophthalmology* 107(5), 453-464.

Radius, R. L. 1978. Perimetry in cataract patients. *Archives of Ophthalmology* 96(9), 1574-1579.

Ransom-Hogg, A. and Spillmann, L. 1980. Perceptive field size in fovea and periphery of the light- and dark-adapted retina. *Vision Research* 20(3), 221-228.

Redmond, T. and Anderson, R. S. 2011. Visual fields: Back to the future. *Optometry in Practice* 12(1), 11-20.

Redmond, T. and Artes, P. H. eds. 2012. *Psychometric functions for visual field stimuli of different size*. 23rd Annual Dalhousie University Ophthalmology Research Meeting. Halifax, Canada.

Redmond, T. et al. 2010a. Sensitivity loss in early glaucoma can be mapped to an enlargement of the area of complete spatial summation. *Investigative Ophthalmology and Visual Science* 51(12), 6540-6548.

Redmond, T. et al. 2010b. The effect of age on the area of complete spatial summation for chromatic and achromatic stimuli. *Investigative Ophthalmology and Visual Science* 51(12), 6533-6539.

Redmond, T. et al. 2011. The effect of stray light on the critical summation area in the peripheral retina. *Investigative Ophthalmology and Visual Science* 52(14), ARVO E-Abstract 5503.

Redmond, T. et al. 2013a. VIsual field progression with frequency doubling matrix perimetry and standard automated perimetry in patients with glaucoma and in healthy controls. *The Journal of the American Medical Association Ophthalmology* 131(12), 1565-1572.

Redmond, T. et al. 2013b. Changes in Ricco's area with background luminance in the S-cone pathway. *Optometry and Vision Science* 90(1), 66-74.

Rempt, F. et al. 1971. Peripheral retinoscopy and the skiagram. *Ophthalmologica* 162(1), 1-10.

Ricco, A. 1877. Relazione fra il minimo angolo visuale e l'intensita luminosa. *Memorie della Regia Academia di Scienze, lettere ed arti in Modena* 17, 47-160.

Richards, W. 1967. Apparent modifiability of receptive fields during accommodation and convergence and a model for size constancy. *Neuropsychologica* 5(1), 63-72.

Rinne, H. 2008. *The Weibull distribution: a handbook*. CRC Press.

Robson, T. 1999. Topics in computerized visual-stimulus generation. In: Carpenter, R.H.S. and Robson, J.G. eds. *Vision Research. A Practical Guide to Laboratory Methods*. Oxford University Press, Oxford, 81-105.

Rountree, L. et al. 2018. Optimising the glaucoma signal/noise ratio by mapping changes in spatial summation with area-modulated perimetric stimuli. *Scientific Reports* 8(1), 2172.

Rowe, F. J. et al. 2013. Comparison of Octopus semi-automated kinetic perimetry and Humphrey peripheral static perimetry in neuro-ophthalmic cases. *ISRN Ophthalmology* 2013, 753202.

Rozema, J. J. et al. 2010. Retinal straylight as a function of age and ocular biometry in healthy eyes. *Investigative Ophthalmology and Visual Science* 51(5), 2795-2799.

Russell, R. A. et al. 2012. The relationship between variability and sensitivity in large-scale longitudinal visual field data. *Investigative Ophthalmology and Visual Science* 53(10), 5985-5990.

Sample, P. A. et al. 2004. Using unsupervised learning with variational Bayesian mixture of factor analysis to identify patterns of glaucomatous visual field defects. *Investigative Ophthalmology and Visual Science* 45(8), 2596-2605.

Sample, P. A. et al. 2006. Identifying glaucomatous vision loss with visual-function-specific perimetry in the diagnostic innovations in glaucoma study. *Investigative Ophthalmology and Visual Science* 47(8), 3381-3389.

Sayed, M. S. et al. 2017. Green disease in optical coherence tomography diagnosis of glaucoma. *Current Opinion in Ophthalmology* 28(2), 139-153.

Schefrin, B. E. et al. 1998. Area of complete scotopic spatial summation enlarges with age. *The Journal of the Optical Society of America A* 15(2), 340-348.

Schefrin, B. E. et al. 1999. Senescent changes in scotopic contrast sensitivity. *Vision research* 39(22), 3728-3736.

Schefrin, B. E. et al. 2004. Evidence against age-related enlargements of ganglion cell receptive field centers under scotopic conditions. *Vision Research* 44(4), 423-428.

Scheifer, U. et al. 2005. Conventional perimetry part 1: Introduction - basic terms. *Der Ophthalmologe* 102(6), 627-646.

Scherk, R. 1872. Ein neuer apparat zur messung des gesichtsfeldes. *Klinische Monatsblatter Fur Augenheilkunde* 10, 151-158.

Scholtes, A. M. W. and Bouman, M. A. 1977. Psychophysical experiments on spatial summation at threshold level of the human peripheral retina. *Vision Research* 17(7), 867-873.

Schulzer, M. 1994. Errors in the diagnosis of visual field progression in normal-tension glaucoma. *Ophthalmology* 101(9), 1589-1594.

Sengpiel, F. and Blakemore, C. 1996. The neural basis of suppression and amblyopia in strabismus. *Eye* 10(2), 250-258.

Shady, S. et al. 2004. Adaptation from invisible flicker. *Proceedings of the National Academy of Sciences of the United States of America* 101(14), 5170-5173.

Shen, C. et al. 2015. Neuroprotective effect of epigallocatechin-3-gallate in a mouse model of chronic glaucoma. *Neuroscience Letters* 600, 132-136.

Shields, M. B. and Spaeth, G. L. 2012. The glaucomatous process and the evolving definition of glaucoma. *Journal of Glaucoma* 21(3), 141-143.

Siddiqui, M. A. et al. 2005. Effect of cataract extraction on frequency doubling technology perimetry in patients with glaucoma. *The British Journal of Ophthalmology* 89(12), 1569-1571.

Silverman, B. W. 1986. *Density estimation for statistics and data analysis*. CRC press.

Simpson, D. A. and Crompton, J. L. 2008a. The visual fields: an interdisciplinary history I. The evolution of knowledge. *Journal of Clinical Neuroscience* 15(2), 101-110.

Simpson, D. A. and Crompton, J. L. 2008b. The visual fields: an interdisciplinary history II. neurosurgeons and quantitative perimetry. *Journal of Clinical Neuroscience* 15(3), 229-236.

Sloan, L. L. 1961. Area and luminance of test object as variables in examination of the visual field by projection perimetry. *Vision Research* 1(1), 121-138.

Sloan, L. L. and Brown, D. J. 1962. Area and luminance of test object as variables in projection perimetry: Clinical studies of photometric dysharmony. *Vision research* 2(12), 527-541.

Smith, D. P. 1925. Centripetal fan scotoma in glaucoma. *The British Journal of Ophthalmology* 9(5), 233-240.

Smith, S. D. et al. 1997. Effect of cataract extraction on the results of automated perimetry in glaucoma. *Archives of Ophthalmology* 115(12), 1515-1519.

Sommer, A. et al. 1984. Evaluation of nerve fiber layer assessment. *Archives of Ophthalmology* 102(12), 1766-1771.

Song, W. et al. 2015. Neuroprotective therapies for glaucoma. *Drug Design, Development and Therapy* 9, 1469-1479.

Sperling, G. 1971. Flicker in computer-generated visual displays: Selecting a CRO phosphor and other problems. *Behavior Research Methods & Instrumentation* 3(3), 151-153.

Spry, P. G. et al. 2000. Quantitative comparison of static perimetric strategies in early glaucoma: test-retest variability. *Journal of Glaucoma* 9(3), 247-253.

Strasburger, H. 2001. Converting between measures of slope of the psychometric function. *Perception and Psychophysics* 63(8), 1348-1355.

Swanson, W. H. 2013. Stimulus size for perimetry in patients with glaucoma. *Investigative Ophthalmology and Visual Science* 54(6), 3984.

Swanson, W. H. et al. 2004. Perimetric defects and ganglion cell damage: Interpreting linear relations using a two-stage neural model. *Investigative Ophthalmology and Visual Science* 45(2), 466-472.

Swanson, W. H. et al. 2005. Quantifying effects of retinal illuminance on frequency doubling perimetry. *Investigative Ophthalmology and Visual Science* 46(1), 235-240.

Swanson, W. H. et al. 2011. Responses of primate retinal ganglion cells to perimetric stimuli. *Investigative Ophthalmology and Visual Science* 52(2), 764-771.

Tafreshi, A. et al. 2009. Visual function-specific perimetry to identify glaucomatous visual loss using three different definitions of visual field abnormality. *Investigative Ophthalmology and Visual Science* 50(3), 1234-1240.

Tanna, A. P. et al. 2004. Impact of cataract on the results of frequency doubling technology perimetry. *Ophthalmology* 111(8), 1504-1507.

Tate, G. W. J. and Lynn, J. R. 1977. *Principles of quantitative perimetry*. New York: Grune and Stratton, Inc.

Thibos, L. N. et al. 1987. Vision beyond the resolution limit: aliasing in the periphery. *Vision Research* 27(12), 2193-2197.

Thibos, L. N. et al. 1996. Characterization of spatial aliasing and contrast sensitivity in peripheral vision. *Vision Research* 36(2), 249-258.

Thomasson, A. H. 1934. A plea for greater uniformity in methods of field taking. *Archives of Ophthalmology* 12(1), 21-32.

Traquair, H. 1927. *An Introduction to Clinical Perimetry*. London: Kimpton.

Traquair, H. M. 1935. Glaucoma, with special reference to medical aspects and early diagnosis. *British Medical Journal* 3906, 933-938.

Traquair, H. M. 1939. Clinical detection of early changes in the visual field. *Transactions of the American Ophthalmological Society* 37, 158-179.

Treutwein, B. and Strasburger, H. 1999. Fitting the psychometric function. *Perception and Psychophysics* 61(1), 87-106.

Trobe, J. D. et al. 1980. An evaluation of the accuracy of community-based perimetry. *American Journal of Ophthalmology* 90(5), 654-660.

Tsatsos, M. and Broadway, D. 2007. Controversies in the history of glaucoma: is it all a load of old Greek? *The British Journal of Ophthalmology* 91(11), 1561-1562.

Turpin, A. 2013. *OPI: Open Perimetry Interface. R package version 1.6*. http://CRAN.R-project.org/package=OPI.

Turpin, A. et al. 2007. Retesting visual fields: utilizing prior information to decrease test-retest variability in glaucoma. *Investigative Ophthalmology and Visual Science* 48(4), 1627-1634.

Turpin, A. et al. 2010. Identifying steep psychometric function slope quickly in clinical applications. *Vision Research* 50(23), 2476-2485.

Turpin, A. et al. 2012. The Open Perimetry Interface: an enabling tool for clinical visual psychophysics. *Journal of Vision* 12(11):22, 1-5.

Tyler, C. W. 1981. Specific deficits of flicker sensitivity in glaucoma and ocular hypertension. *Investigative Ophthalmology and Visual Science* 20(2), 204-212.

Ueda, T. et al. 2006. Frequency doubling technology perimetry after clear and yellow intraocular lens implantation. *American Journal of Ophthalmology* 142(5), 856-858.

Urcola, J. H. et al. 2006. Three experimental glaucoma models in rats: Comparison of the effects of intraocular pressure elevation on retinal ganglion cell size and death. *Experimental Eye Research* 83(2), 429-437.

Van den Berg, T. J. 1986. Importance of pathological intraocular light scatter for visual disability. *Documenta Ophthalmologica* 61(3-4), 327-333.

Van den Berg, T. J. 1995. Analysis of intraocular straylight, especially in relation to age. *Optometry and Vision Science* 72(2), 52-59.

Van den Berg, T. 2017. The (lack of) relation between straylight and visual acuity. Two domains of the point-spread-function. *Ophthalmic and Physiological Optics* 37(3), 333-341.

Van den Berg, T. J. et al. 1991. Dependence of intraocular straylight on pigmentation and light transmission through the ocular wall. *Vision Research* 31(7-8), 1361-1367.

Van den Berg, T. J. et al. 2005. New approach for retinal straylight assessment: compensation comparison. *Investigative Ophthalmology and Visual Science* 46(13), ARVO E-Abstract 4315.

Van den Berg, T. J. et al. 2007. Straylight effects with aging and lens extraction. *American Journal of Ophthalmology* 144(3), 358-363.

Vassilev, A. et al. 2000. Spatial summation of blue-on-yellow light increments and decrements in human vision. *Vision Research* 40(8), 989-1000.

Vassilev, A. et al. 2003. Spatial summation of S-cone ON and OFF signals: Effects of retinal eccentricity. *Vision Research* 43(27), 2875-2884.

Vassilev, A. et al. 2005. Human S-cone vision: relationship between perceptive field and ganglion cell dendritic field. *Journal of Vision* 5(10), 823-833.

Viswanathan, A. C. et al. 1997. How often do patients need visual field tests? *Graefe's Archive for Clinical and Experimental Ophthalmology* 235(9), 563-568.

Volbrecht, V. J. et al. 2000a. Ricco's area for S- and L-cone mechanisms across the retina. *Color Research and Application* 26(51), S32-S35.

Volbrecht, V. J. et al. 2000b. Spatial summation in human cone mechanisms from 0 to 20 degrees in the superior retina. *Journal of the Optical Society of America. A* 17(3), 641-650.

Von Graefe, A. 1856. Ueber die untersuchung des gesichtsfeldes bei amblyopischen affectionen. *Archiv für Ophthalmologie* 2(2), 258-298.

Wall, M. et al. 1991. Variability of high-pass resolution perimetry in normals and patients with idiopathic intracranial hypertension. *Investigative Ophthalmology and Visual Science* 32(12), 3091-3095.

Wall, M. et al. 1996. The psychometric function and reaction times of automated perimetry in normal and abnormal areas of the visual field in patients with glaucoma. *Investigative Ophthalmology and Visual Science* 37(5), 878-885.

Wall, M. et al. 1997. Variability in patients with glaucomatous visual field damage is reduced using size V stimuli. *Investigative Ophthalmology and Visual Science* 38(2), 426-435.

Wall, M. et al. 2004. The effect of attention on Conventional Automated Perimetry and Luminance Size Threshold Perimetry. *Investigative Ophthalmology & Visual Science* 45(1), pp. 342-350.

Wall, M. et al. 2009. Repeatability of automated perimetry: a comparison between standard automated perimetry with stimulus size III and V, matrix, and motion perimetry. *Investigative Ophthalmology and Visual Science* 50(2), 974-979.

Wall, M. et al. 2013. Size threshold perimetry performs as well as conventional automated perimetry with stimulus sizes III, V, and VI for glaucomatous loss. *Investigative Ophthalmology and Visual Science* 54(6), 3975-3983.

Wang, Y. and Henson, D. B. 2013. Diagnostic performance of visual field test using subsets of the 24-2 test pattern for early glaucomatous field loss. *Investigative Ophthalmology and Visual Science* 54(1), 756-761.

Wang, Y.Z. et al. 1997. Effects of refractive error on detection acuity and resolution acuity in peripheral vision. *Investigative Ophthalmology and Visual Science* 38(10), 2134-2143.

Watson, A. B. and Pelli, D. G. 1983. QUEST: a Bayesian adaptive psychometric method. *Perception and Psychophysics* 33(2), 113-120.

Weinreb, R. N. et al. 2014. The pathophysiology and treatment of glaucoma: a review. *The Journal of the American Medical Association* 311(18), 1901-1911.

Werner, E. B. et al. 1988. Effect of patient experience on the results of automated perimetry in clinically stable glaucoma patients. *Ophthalmology* 95(6), 764-767.

Wetherill, G. B. and Levitt, H. 1965. Sequential estimation of points on a psychometric function. *British Journal of Mathematical and Statistical Psychology* 18(1), 1-10.

White, A. J. R. et al. 2002. An examination of physiological mechanisms underlying the frequency doubling illusion. *Investigative Ophthalmology and Visual Science* 43(11), 3590-3599.

Wichmann, F. A. and Hill, N. J. 2001a. The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception and Psychophysics* 63(8), 1293-1313.

Wichmann, F. and Hill, N. J. 2001b. The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception and Psychophysics* 63(8), 1314-1329.

Wiesel, T. N. 1960. Receptive fields of ganglion cells in the cat's retina. *The Journal of Physiology* 153(3), 583-594.

Wild, J. M. 2001. Short wavelength automated perimetry. *Acta Ophthalmologica Scandinavica* 79(6), 546-559.

Wild, J. M. and Moss, I. D. 1996. Baseline alterations in blue-on-yellow normal perimetric sensitivity. *Graefe's Archive for Clinical and Experimental Ophthalmology* 234(3), 141-149.

Wild, J. M. et al. 1989. The influence of the learning effect on automated perimetry in patients with suspected glaucoma. *Acta Ophthalmologica (Copenhagen)* 67(5), 537-545.

Wild, J. M. et al. 1991. Long-term follow-up of baseline learning and fatigue effects in the automated perimetry of glaucoma and ocular hypertensive patients. *Acta Ophthalmologica* 69(2), 210-216.

Wild, J. M. et al. 1998. Statistical aspects of the normal visual field in short-wavelength automated perimetry. *Investigative Ophthalmology and Visual Science* 39(1), 54-63.

Wilensky, J. T. and Joondeph, B. C. 1984. Variation in visual field measurements with an automated perimeter. *American Journal of Ophthalmology* 97(3), 328-331.

Wilensky, J. T. et al. 1986. The use of different-sized stimuli in automated perimetry. *American Journal of Ophthalmology* 101(6), 710-713.

Wilson, M. E. 1970. Invariant features of spatial summation with changing locus in the visual field. *The Journal of Physiology* 207(3), 611-622.

Wood, J. M. et al. 1989. Alterations in the shape of the automated perimetric profile arising from cataract. *Graefe's Archive for Clinical and Experimental Ophthalmology* 227(2), 157-161.

Yücel, Y. H. et al. 2000. Loss of neurons in magnocellular and parvocellular layers of the lateral geniculate nucleus in glaucoma. *Archives of Ophthalmology* 118(3), 378-384.

Zeiss. 2014. Humphrey visual field analyzers, innovation and connectivity [Brochure]. *Carl Zeiss Meditec*.

Zeiss. 2015. Humphrey Field Analyzer 3 from Zeiss; The best just got better [Brochure]. *Carl Zeiss Meditec.*

Zele, A. J. and Vingrys, A. J. 2005. Cathode-ray-tube monitor artefacts in neurophysiology. *Journal of Neuroscience Methods* 141(1), 1-7.

Zlatkova, M. B. et al. 2006. The effect of simulated lens yellowing and opacification on blue-on-yellow acuity and contrast sensitivity. *Vision Research* 46(15), 2432-2442.

Zychaluk, K. and Foster, D. H. 2009. Model-free estimation of the psychometric function. *Attention, Perception, and Psychophysics* 71(6), 1414-1425.

Appendix A – Example FOS curves

Example FOS curves from data collected in chapter four, fitted with a logistic psychometric function using MATLAB (version R2015b; The MathWorks Inc., Natick, MA, USA), and the Palamedes toolbox (Prins and Kingdom 2009).

(A) Complete FOS curves (healthy) for each of the four stimulus forms, at each of the four test locations.
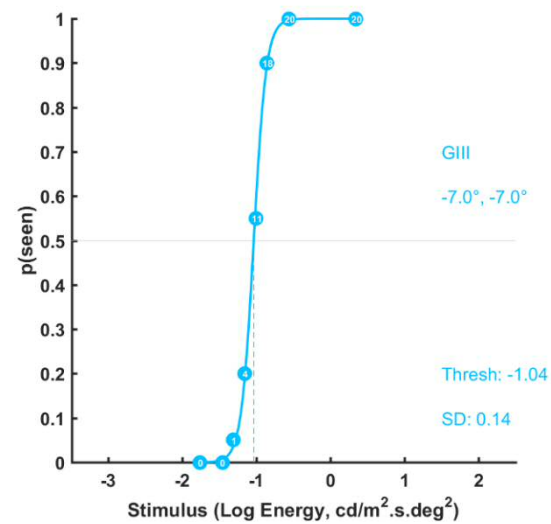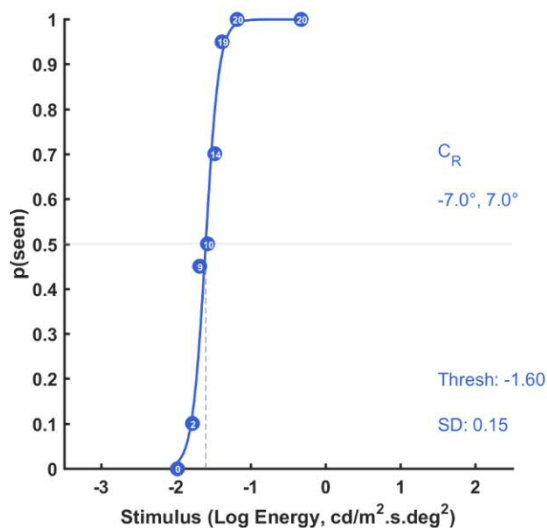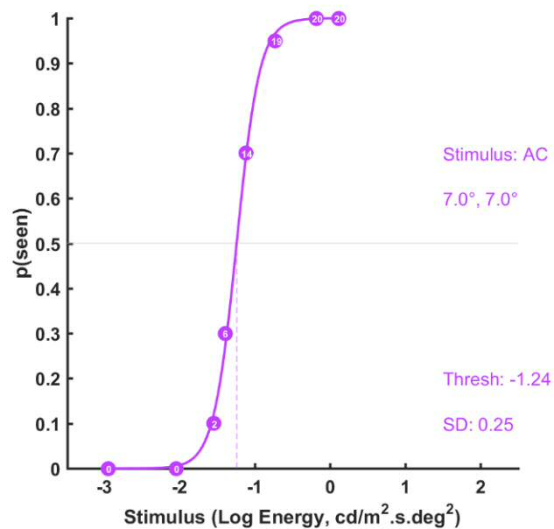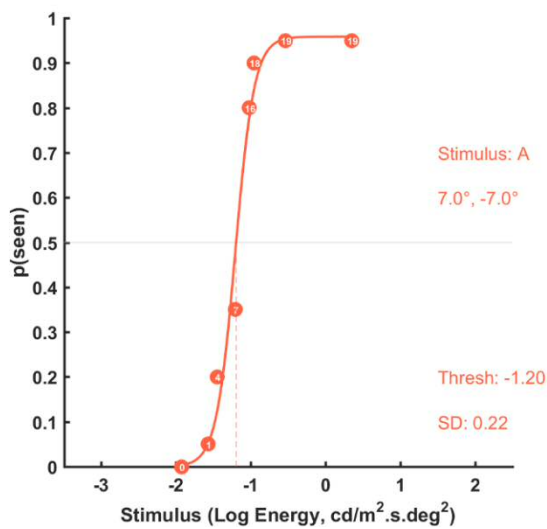
(B) Complete FOS curves (glaucoma) for each of the four stimulus forms, at each of the four test locations.
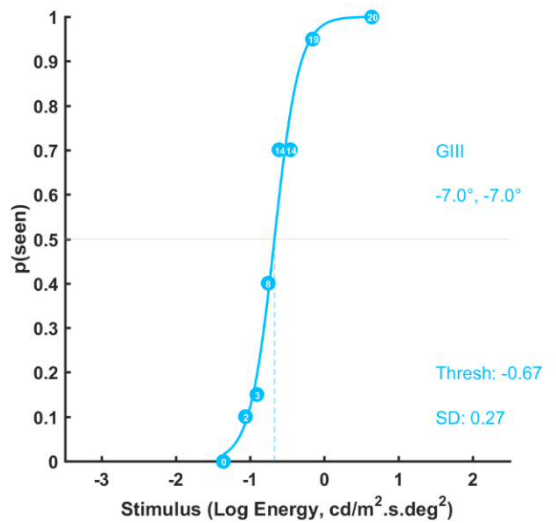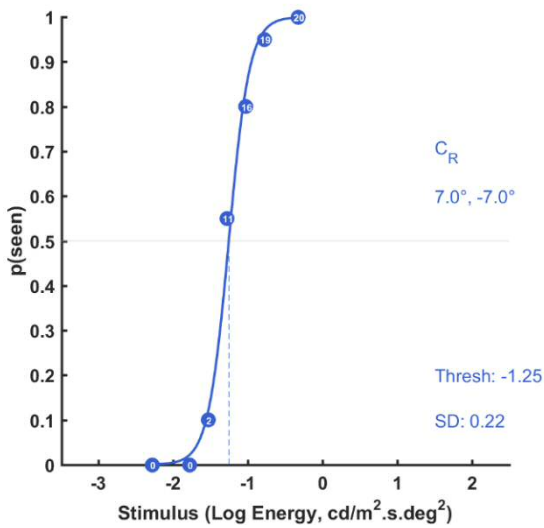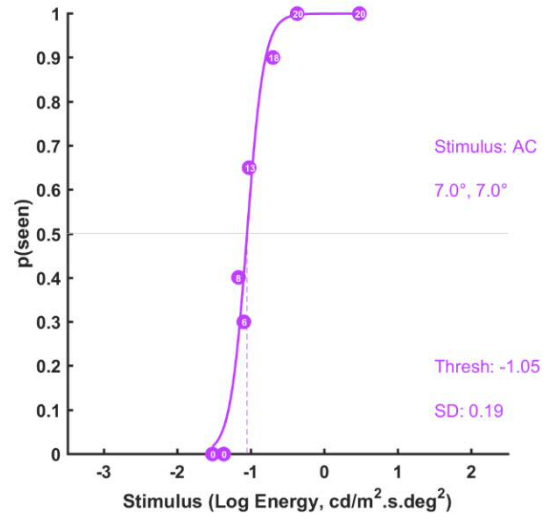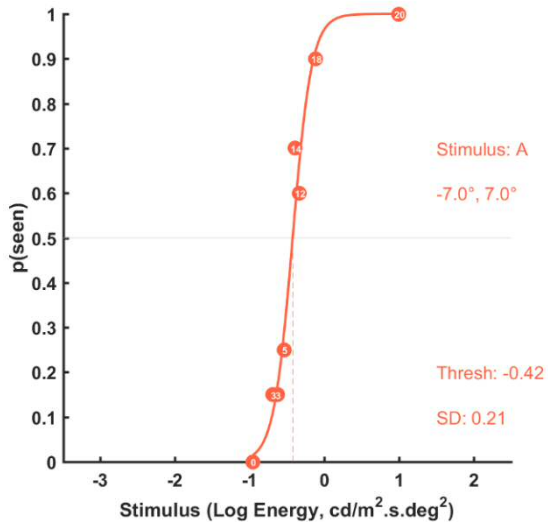
(C) Incomplete FOS curves (glaucoma) for each of the four stimulus forms, at each of the four test locations.

Complete FOS curves (p(seen) = 1.0) were not attainable in these examples due to severity of disease. In those FOS curves where the observer did not respond to any stimulus value at least half the time, a threshold estimate would still be provided, defaulting to the maximum stimulus value available for that stimulus (example shown for the C_R stimulus). These threshold estimates were excluded.
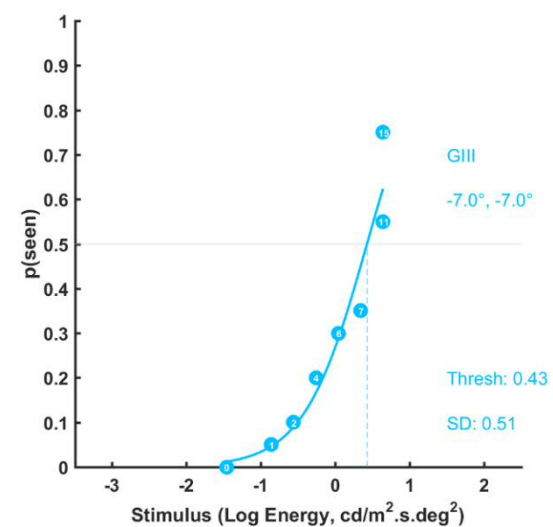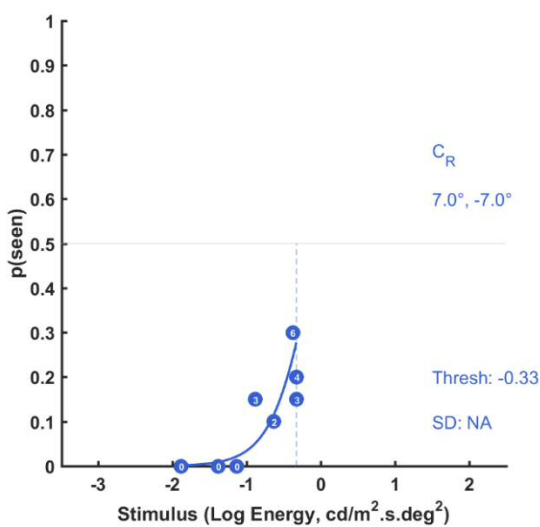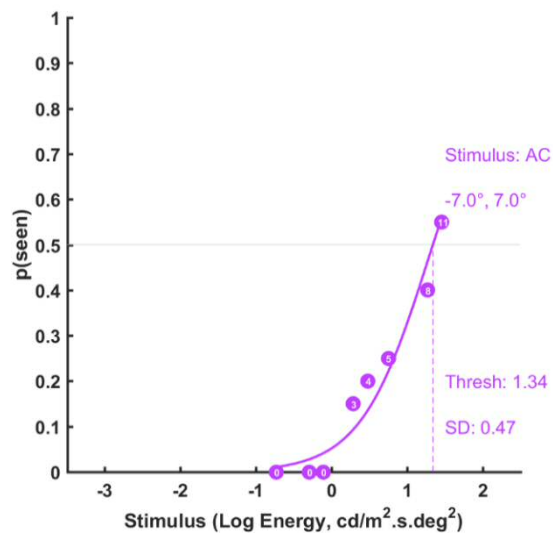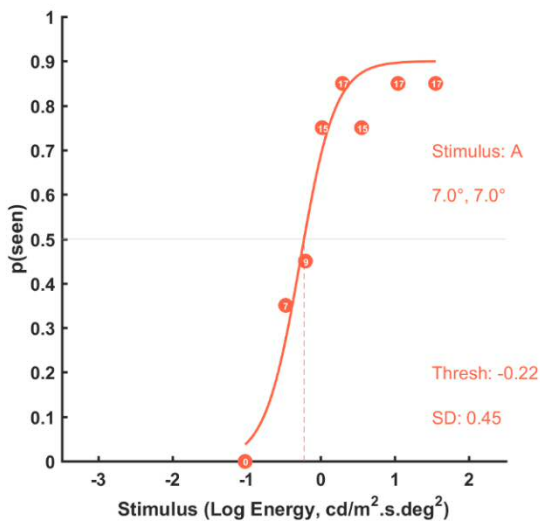
Appendix B – Example participant information leaflet

PARTICIPANT INFORMATION SHEET
(Phase (i): Patients with Glaucoma)

Multi-dimensional stimuli for clinical perimetry

What is this study about?

The visual field test is commonly used to investigate glaucoma, both in an optometrist's practice and in the hospital eye service. This normally involves sitting at a machine, looking at a central spot of light, and pressing a button whenever another small spot of light appears in the peripheral (side) vision. While this instrument has been relied upon for many years to aid in the diagnosis and monitoring of glaucoma, the results can be somewhat variable. As such, multiple visual field tests (generally over the course of weeks or months) are required in order to arrive at a confident conclusion.

We are conducting a study that aims to minimise the variability of the visual field test, and as such make it more effective at identifying very early glaucoma, and identifying when established glaucoma is getting worse. In the current visual field test, the spots of light vary in brightness, while the size of the spots stay the same. We want to investigate whether varying the spots in a different way will be more effective, and will investigate 4 methods:

- The spots varying in brightness, while the size of the spots stay the same (the spots will be the same size as those used in the current visual field test)
- The spots varying in brightness, while the size of the spots stay the same (the spots will be smaller than those used in the current visual field test)
- The spots varying in size, while the brightness of the spots stay the same
- The spots varying in both size and brightness simultaneously

We will then compare the variability between these tests.

This study is being carried out as part of a degree of Doctor of Philosophy (PhD) by Lindsay Rountree at Cardiff University.

This is the first phase of the study. In this phase, we are testing which of the methods described above work the best. We will also be conducting the same tests with a group of people who do not have glaucoma. We will then be able to compare the results from each of the 4 tests to see which gives us the most accurate, and least variable, assessment of visual field.

Why am I being asked to join the study?

You have been asked to take part in the study because we are looking for a number of volunteers who have either Primary Open Angle Glaucoma (POAG) or Normal Tension Glaucoma (NTG). If you are unsure about the current health of your eyes, we recommend that you visit your optometrist or ophthalmologist for a check-up.

What does the study involve?

The tests are similar to a standard visual field test. We will ask you to keep looking at a constant spot of light in the centre of a screen, and press a button every time you see another light flash in your peripheral (side) vision. We will only test one of your eyes. As described above, there will be 4 different tests. Unlike the standard visual field test, we will test fewer parts of your visual field, but we will test them much more thoroughly.

Each test will take approximately 45 minutes, with regular rest periods. You may also request rest periods at any time if you feel tired. As these tests are time consuming and can be tiring, the 4 different tests will be carried out on 4 different days. We will also undertake some preliminary clinical measurements in order to confirm your eligibility for the study, which will be described below. These will take approximately 45 minutes in total. The total test time will therefore be approximately 1.5 hours on day 1, and 1 hour on days 2, 3 and 4. The days do not have to be consecutive, and can be carried out at your convenience.

How will I know if I am suitable?

We are looking for people who have no other eye conditions apart from glaucoma. You also need to have good vision on a standard letter chart, with or without spectacles/contact lenses. Many people with advanced glaucoma are still able to achieve this. Your spectacle prescription needs to be within +6.00 and -6.00, with less than -3.00 astigmatism (you may ask your optometrist or the investigators for advice if you are unsure about your spectacle prescription). Unfortunately, people with conditions such as diabetes or thyroid disease are not eligible for inclusion in this study. If you are unsure if your eyes are healthy, we recommend that you visit your GP or optometrist for advice, or a check-up if this is due.

To confirm that you are eligible to take part in the study, you will be required to undergo some preliminary tests, most of which you will have had before during a routine visit to your optometrist. This will include:
- Measuring your vision on a letter chart and checking your spectacle prescription
- Measuring the pressure inside your eye
- A standard visual field test (as is commonly used in hospitals and optometric practice)
- A check of the health of your eye
- A scan of your retina (the layer at the back of the eye) with the Heidelberg Retina Tomograph 3. This is a commercially available instrument, and does not come into contact with the eye; it feels similar to having a photograph taken.

Please note that for the pressure measurement, we will need to instil 1 drop of a mild anaesthetic eye drop (proxymetacaine 0.5% or benoxinate 0.4%). This is the normal clinical method of checking the pressure in your eye during a routine visit to the optometrist or ophthalmologist. This drop will not dilate the pupil of your eye, nor cause the vision to become blurred.

If you have had a visual field test in the past 6 months, and you (or University Hospital of Wales/Cardiff University) have a copy of the results, you may only have to undergo 1 of these tests as part of this study. However, if you have not had a visual field test in the past 6 months, you will be required to undergo 2 visual field tests. (Note: These visual field tests are separate to those carried out as part of the main experiment. These tests are to confirm your eligibility only).

You will also be asked about any eye-related condition or medication that may affect your vision. Similarly, some general health conditions or medications can affect vision; you will be asked about any such medication before you are deemed suitable for inclusion in the study.

Please note that the tests of your vision carried out during this research do not constitute a regular "sight test" as carried out by a registered optometrist. Although the results may be useful to you, they are not a substitute for regular visits to your optometrist or ophthalmologist.

Participants should ***not*** refrain from taking any regular medication (if applicable) due to this study. Please note that we are ***not*** investigating a new treatment for glaucoma, so any medication that you are currently taking must be continued unless advised otherwise by your doctor.

Contact lens wearers are asked not to wear their lenses for the following periods prior to the tests:
- Soft lens wearers: ***3 hours***
- Rigid Gas Permeable (RGP) lens wearers: ***6 hours***

If you encounter any circumstance which means you may no longer be eligible for inclusion in this project, we would be grateful if you would speak to a member of the research team.

The investigators will be able to offer £10 to compensate you for the time you have given up for the research.

Are there any side effects?

There are no known side effects from any of the experimental tests.

The mild anaesthetic eye drop (used for the pressure measurement) will wear off after approximately 20 minutes. You are advised not to rub your eye during this time.

Sometimes, an abnormality of the eye may be discovered which the participant had previously been unaware of. In the event of any abnormality of the eye being detected during the course of this study, you will be referred to your GP, with your consent, in the same manner as you would be referred by your local optometrist.

What happens if something goes wrong?

It is very unlikely that anything will go wrong, as the tasks are similar to watching TV. However, you will have immediate access to an eye care practitioner at Cardiff University in the very unlikely event that something adverse happens.

How long will I have to be in the clinic?

The preliminary tests will take approximately 45 minutes (or 55 minutes, depending on whether you will need to undergo 1 or 2 standard visual field tests). Each experimental test will take approximately 45 minutes, including breaks. Therefore, it is likely that you will be in the clinic for approximately 1.5 hours on day 1 and 1 hour on days 2, 3 and 4. You may take rests and can have refreshments at any point. There are rest periods timed during and after each test, and you may request a break at any time. Of course, if you would like more breaks, the overall time of the experiment may take a little longer.

Benefits

There are no direct benefits for you, except the knowledge that you are helping with the continued development of tests aimed at monitoring glaucoma progression.

Duration of the study

All the tests are performed during 4 visits. You may change your mind about taking part in the study at any time. Withdrawal from the study will not affect the level of care that you currently receive (or will receive in the future) as part of your visits to the optometrist or any NHS institution.

Confidentiality

Strict confidentiality will be upheld at all times. Personal information will be kept securely under the terms of the Data Protection Act (1998) during the course of the study. Data collected will be used as part of a PhD study; all data will be anonymised before any analysis is carried out. On completion of the study, data may be kept for use in future studies, but it will remain anonymised. With your consent, your GP will be informed of your participation in this research.

What will happen to the results of this study?

The results will be written up as a report, which will be published as a research article in a leading research journal. They will also be presented at a research conference on completion of the project. All data will be anonymised; no individual will be identifiable from his/her data.

A summary of the results will be available from the Principal Investigator following completion of the project, on request.

Voluntary participation

Your participation in this study is entirely voluntary, and should you prefer not to participate, your routine clinical care will not be affected in any way. Please note that you can withdraw from the study at any time. You may invite someone along to accompany you to the clinic, if you so wish.

Who is organising and funding the research?

This research study is organised and sponsored by Cardiff University. It is funded by the College of Optometrists.

Who has reviewed this study?

Every research investigation is looked at by an independent group of people, called a Research Ethics Committee, to protect your safety, rights, wellbeing and dignity. The East of Scotland Research Ethics Service REC 2, which has responsibility for scrutinising all proposals for medical research on humans in Tayside, has examined the proposal and has raised no objections from the point of view of medical ethics (Ref: 15/ES/0070). It is a requirement that your records in this research, together with any relevant medical records, be made available for scrutiny by monitors from Cardiff University and NHS Cardiff and Vale Health Board, whose role is to check that research is properly conducted and the interests of those taking part are adequately protected.

Where can I find independent advice on participating in the research study?

Prof. Rachel North at the School of Optometry and Vision Sciences, Cardiff University is familiar with this study, although she is not directly involved. She may be contacted at:

The School of Optometry and Vision Sciences, Cardiff University
Maindy Road
Cardiff, CF24 4HQ
Tel: 029 2087 5114

If you would like independent advice on participating in research studies in general, we recommend that you contact the International Glaucoma Association (IGA, www.glaucoma-association.com) or the College of Optometrists (Research section, www.college-optometrists.org/en/research/index.cfm)

Complaints

The research does not carry any more risks than visiting the hospital or optometrist in the normal way. If taking part in this research project harms you, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for a legal action against Cardiff University or the NHS Health Board but you may have to pay for your legal costs. The normal National Health Service complaints mechanisms will still be available to you.

If you wish to make a complaint, you can raise this with the Research Team (contact details of the Principal Investigator, Dr Tony Redmond, can be found below). If you wish to complain to someone outside of the Research Team, then you have two options depending on whether the complaint relates to activities at a NHS or non-NHS site:

Dr Tony Redmond (Principal Investigator) Tel: 029 2087 0564 Email: RedmondT1@cardiff.ac.uk	Miss Lindsay Rountree (Investigator) Tel: 029 2087 0247 Email: RountreeLC@cardiff.ac.uk
Prof. James E. Morgan Tel: 029 2087 6344 Email: MorganJE3@cardiff.ac.uk	School of Optometry and Vision Sciences Cardiff University Maindy Road Cardiff, CF24 4HQ

Appendix C – Example consent form

PARTICIPANT INFORMATION SHEET
(Phase (i): Patients with Glaucoma)

Multi-dimensional stimuli for clinical perimetry

Approval No: 15/ES/0070
Version: 2.0 (29th April 2015)

Name of Researchers: **Dr Tony Redmond**
Miss Lindsay Rountree
Prof. Roger Anderson
Dr Pádraig Mulholland
Prof. James E. Morgan

Please initial box

1   I confirm that I have read and understand the Participant Information Sheet for Phase (i): Patients with Glaucoma (Version: 2.0, Date: 29th April 2015) for the above study and have had the opportunity to ask questions

2   I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason

3   I consent to my GP being informed of my participation in this research

4   I understand that relevant sections of my medical notes, and data collected during the study, may be looked at by individuals from Cardiff University or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records

5   I agree to take part in the above study

Name of participant   Date   Signature

Researcher   Date   Signature

Name of person taking consent   Date   Signature
(if different from researcher)

Appendix D – Ethical approval

East of Scotland Research Ethics Service *(EoSRES)*

NHS SCOTLAND

Research Ethics Service

TAyside medical Science Centre
Residency Block Level 3
George Pirie Way
Ninewells Hospital and Medical School
Dundee DD1 9SY

Dr Tony Redmond
Lecturer & Deputy Director of Postgraduate Research
Cardiff University
School of Optometry and Vision Sciences, Cardiff
University
Maindy Road
Cardiff, CF24 4HQ

Date: 18 May 2015
Your Ref:
Our Ref: **DL/15/ES/0070**
Enquiries to: Mrs Diane Leonard
Direct Line: 01382 383871
Email: eosres.tayside@nhs.net

Dear Dr Redmond

Study title: **Functional Mapping of the Biological Changes in the Retina in Glaucoma**
REC reference: 15/ES/0070
Protocol number: SPON 1406-15
IRAS project ID: 152375

Thank you for your letter of 29 April 2015, responding to the Proportionate Review Sub-Committee's request for changes to the documentation for the above study.

The revised documentation has been reviewed and approved by the sub-committee.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact the Assistant Co-ordinator, Mrs Diane Leonard, eosres.tayside@nhs.net. Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Conditions of the favourable opinion

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

1

Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of approvals from host organisations.

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.
There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from NRES. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" above).

Approved documents

The documents reviewed and approved by the Committee are:

Document	Version	Date
Copies of advertisement materials for research participants	v2	29 April 2015
Covering letter on headed paper		29 April 2015
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only)		16 July 2014
GP/consultant information sheets or letters	v2	29 April 2015
IRAS Checklist XML [Checklist_12052015]		12 May 2015

2

Document	Version	Date
Letter from funder		19 December 2013
Letter from sponsor		01 April 2015
Letters of invitation to participant	1.0	17 March 2015
Letters of invitation to participant	1.0	17 March 2015
Letters of invitation to participant	1.0	17 March 2015
Letters of invitation to participant	1.0	17 March 2015
Other [CV - second supervisor - Professor Roger S. Anderson]		
Other [Summary CV - Third supervisor - PÃ¡draig Mulholland]		
Other [Summary CV - second student - Shindy Je]		
Participant consent form [Phase (iii)]	1.0	17 March 2015
Participant consent form [(Phase (i): Patients with Glaucoma)]	2	29 April 2015
Participant consent form [(Phase (i): Healthy Participants)]	2	29 April 2015
Participant consent form [(Phase (ii): Patients with Glaucoma)]	2	29 April 2015
Participant consent form [(Phase (ii): Healthy Participants)]	2	29 April 2015
Participant consent form [Phase (iii)]	2	29 April 2015
Participant information sheet (PIS) [(Phase (i): Patients with Glaucoma)]	2	29 April 2015
Participant information sheet (PIS) [(Phase (i): Healthy Participants)]	2	29 April 2015
Participant information sheet (PIS) [(Phase (ii): Patients with Glaucoma)]	2	29 April 2015
Participant information sheet (PIS) [(Phase (ii): Healthy Participants)]	2	29 April 2015
Participant information sheet (PIS) [Phase (iii)]	2	29 April 2015
REC Application Form [REC_Form_13042015]		13 April 2015
Research protocol or project proposal	1.0	17 March 2015
Summary CV for Chief Investigator (CI) [Dr Tony Redmond]		
Summary CV for student [Lindsay Rountree]		
Summary CV for supervisor (student research) [Dr Tony Redmond]		

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators

3

☐ Notification of serious breaches of the protocol
☐ Progress and safety reports
☐ Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

Feedback

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website:
http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

15/ES/0070	Please quote this number on all correspondence

Yours sincerely

D Leonard.

for Dr Anthony Davis
Vice-Chair

E-mail: eosres.tayside@nhs.net

Enclosures: "After ethical review – guidance for researchers"

Copy to: Miss Lindsay Rountree
                      Ms Emma Lewis, R&D Cardiff and Vale University Health Board

4

Appendix E – Curriculum vitae

LINDSAY ROUNTREE
BSc (Hons) MCOptom

School of Optometry and Vision Sciences, Cardiff University
General Optical Council, Reg. no: 01-23415

EDUCATION

Jan 2014-Dec 2017	**School of Optometry and Vision Sciences, Cardiff University, Cardiff** PhD. Thesis submitted 19th December
Apr 2012	**WOPEC, School of Optometry and Vision Sciences, Cardiff University, Cardiff** LOCSU Glaucoma Level 1 and 2
Jan 2011	**Department of Optometry, City University, London** MSc Module in Binocular Vision
Oct 2008	**Department of Optometry, City University, London** MSc Module in Optometric Management of Glaucoma
Nov 2007	**Kent Eye Screening Service, Ashford** Accredited in Diabetic Screening by Slit Lamp BIO
Aug 2006-Jul 2007	**Scheme for Registration, College of Optometrists, London** Completed and was subsequently accepted into the College of Optometrists
Sept 2003-Jul 2006	**School of Optometry and Vision Sciences, Cardiff University, Cardiff** 2.1 BSc (Hons) Degree in Optometry and Vision Sciences

ACADEMIC EXPERIENCE

Jan 2014-Dec 2017	**PhD student, School of Optometry and Vision Sciences, Cardiff University, Cardiff** Thesis: Optimisation of perimetric stimuli for mapping changes in spatial summation in glaucoma. Supervisors: Dr Tony Redmond, Prof. Roger S. Anderson and Dr Pádraig J. Mulholland.
Jan 2014-Present	**Undergraduate Clinic Supervisor, School of Optometry and Vision Sciences, Cardiff University, Cardiff** Regular supervision of a range of undergraduate modules and year groups.

Oct 2015-May 2016	**Co-supervisor, School of Optometry and Vision Sciences, Cardiff University, Cardiff**
	Co-supervision of 3rd year optometry undergraduate students during their final year research project.
Mar 2016-Dec 2016	**Optometrist, Clinical Trials, Cardiff and Vale UHB**
	Refraction of patients participating in clinical trials.

OTHER EXPERIENCE

	Optometrist working in practice at the following locations:
Jan 2014 - Present	Locum optometrist, regularly working at Bath Opticians, Bath since August 2014
Jun 2011-Dec 2014	Full-time optometrist at R. J. Holmes Opticians, Cirencester
Feb 2011-Jun 2011	Full-time optometrist at The Milsom Eye Company, Bath
Oct 2009-Feb 2011	Volunteer optometrist on days off, with Vision Care for Homeless People, London
Aug 2007-Feb 2011	Full-time optometrist at Leslie Warren Opticians/Linklaters Optometrists, Sevenoaks/Bexleyheath
Aug 2006-Aug 2007	Pre-registration optometrist at Leslie Warren Opticians/Linklaters Optometrists, Sevenoaks/Bexleyheath

CONFERENCE PRESENTATIONS

Jun 2017	**Rountree L**, Mulholland PJ, Anderson RS, Morgan JE, Garway-Heath DF, Redmond T. Quantifying the signal/noise ratio with perimetric stimuli optimised to probe changing spatial summation in glaucoma. *Vision Researchers' Colloquium*. University of Bath, Bath, UK. *(Poster presentation)*
May 2017	**Rountree L**, Mulholland PJ, Anderson RS, Morgan JE, Garway-Heath DF, Redmond T. Quantifying the signal/noise ratio with perimetric stimuli optimised to probe changing spatial summation in glaucoma. *Investigative Ophthalmology and Visual Science*; 58: ARVO E-Abstract 2852. Baltimore, ML, USA. *(Poster presentation)*
May 2017	Redmond T, **Rountree L**, Anderson RS, Mulholland PJ. Resilience of area-modulated perimetric stimuli to increased intraocular straylight. *Investigative Ophthalmology and Visual Science*; 58: ARVO E-Abstract 2856. Baltimore, ML, USA. *(Poster presentation)*
Mar 2017	**Rountree L**, Mulholland PJ, Anderson RS, Morgan JE, Garway-Heath DF, Redmond T. Modulating area, rather than intensity, in perimetric stimuli may increase the disease signal in glaucoma. *Optometry Tomorrow*. Birmingham, UK. *(Poster presentation)*

Mar 2017	Esterer S, **Rountree L,** Redmond T, MᶜGonigle DJ. Plan for the worst, hope for the best: towards individualising stimulation and experimental parameters in tDCS. *Brain Stimulation*; 10(2): 477-478.
Sep 2016	**Rountree L**, Mulholland PJ, Anderson RS, Redmond T. Response variability for multi-dimensional perimetric stimuli in glaucoma. *Imaging & Perimetry Society (IPS) 22nd International Visual Field & Imaging Symposium*. Udine, Italy *(Paper presentation)*
Sep 2016	**Rountree L**, Mulholland PJ, Anderson RS, Redmond T. Response variability for multi-dimensional perimetric stimuli in glaucoma. *Ophthalmic and Physiological Optics*; 36(6): 8. *(Paper presentation)*
May 2015	**Rountree L**, Mulholland PJ, Anderson RS, Redmond T. Response variability for perimetric stimuli of different area in healthy observers. *European Academy of Optometry and Optics (EAOO)*. Budapest, Hungary. *(Paper presentation)*
Mar 2015	**Rountree L** and Redmond T. Response variability across the visual field with perimetric stimuli of different area. *Optometry Tomorrow*. Brighton, England, UK. *(Poster presentation; CET accredited)*
Sep 2014	**Rountree L** and Redmond T. Response variability across the visual field with perimetric stimuli of different area. *Ophthalmic and Physiological Optics*; 34(6): 684. *(Paper presentation)*
May 2014	**Rountree L**. Does size matter? Improving glaucoma management by changing how we measure peripheral vision. *Speaking of Science*. Cardiff University, Cardiff, Wales, UK. *(Paper presentation)*

INVITED LECTURES

Mar 2016	**Rountree L**. Visual field testing and its place in modern practice. *100% Optical*. London, England, UK *(Lecture presentation; CET accredited)*

PUBLICATIONS

Feb 2018	**Rountree L**, Mulholland PJ, Anderson RS, Garway-Heath DF, Morgan JE, Redmond T. Optimising the glaucoma signal/noise ratio by mapping changes in spatial summation with area-modulated perimetric stimuli. *Scientific Reports* 8(1), 2172.

ARTICLES

Aug 2015   **Rountree L** and Redmond T. Visual fields: where are we now?
Optometry Today. Aug 2015; 48-52.
(CET Article)

SELECTED HONOURS AND AWARDS

Feb 2017   **College of Optometrists Research Travel Bursary**
For attendance at the Annual Association for Research in Vision and
Ophthalmology (ARVO), Baltimore, USA

Jul 2016   **Imaging and Perimetry Society (IPS) Travel Bursary**
For attendance at IPS, Udine, Italy

Sept 2016   **British Congress of Optometry and Vision Science (BCOVS)**
University of Ulster, Coleraine, Northern Ireland, UK
2nd prize, paper presentation

Mar 2015   **Optometry Tomorrow**
Brighton, England, UK
Award for poster presentations

May 2014   **Speaking of Science**
Cardiff University, Cardiff, Wales, UK
2nd prize, paper presentation