Insights for mfVEPs from perimetry using large spatial frequency doubling and near frequency doubling stimuli in glaucoma

Siti Nurliyana Abdullah[1,2], Gordon F Sanderson[3,†], Mohd Aziz Husni[4], Ted Maddess[2]

[1]Orthoptic Unit, Eye Centre, RIPAS Hospital, Jalan Putera Al-Muhtadee Billah, Bandar Seri Begawan BA 1710, Brunei Darussalam

[2]Eccles Institute for Neuroscience, John Curtin School of Medical Research, Australian National University, Canberra, Australia

[3]Ophthalmology Section, Department of Medicine, Otago University, Dunedin, New Zealand

[4]Department of Ophthalmology, Hospital Selayang, 68100 Batu Caves, Selangor Darul Ehsan, Malaysia

†Deceased

Address for all correspondence:
Prof T Maddess
Eccles Institute for Neuroscience
John Curtin School of Medical Research (Bldg 131)
Australian National University
Canberra ACT 2601, Australia
Tel. +61 2 6125 4099
Fax. +61 2 6125 9532
ted.maddess@anu.edu.au

ORCID: Abdullah 0000-0001-7230-4618, Aziz 0000-0001-9949-364x, Maddess 0000-0003-4591-3658

Words: 3839

2 Tables, 5 Figures

Acknowledgement: This paper is dedicated to our late colleague Gordon F Sanderson. We are grateful for the very constructive comments of the reviewers.

Abstract (349 words)

Purpose To compare two forms of perimetry that use large contrast-modulated grating stimuli in terms of: their relative diagnostic power, their independent diagnostic information about glaucoma, and their utility for mfVEPs. We evaluated a contrast-threshold mfVEP in normal controls using the same stimuli as one of the tests.

Methods We measured psychophysical contrast-thresholds in one eye of 16 control subjects and 19 patients aged 67.8 ± 5.65 and 71.9 ± 7.15 respectively (mean ± SD). Patients ranged in disease severity from suspects to severe glaucoma. We used the 17-region FDT-perimeter C20-threshold program and a custom 9-region test (R9) with similar visual field coverage. The R9 stimuli scaled their spatial frequencies with eccentricity and were modulated at lower temporal frequencies than C20, and thus did not display a clear spatial frequency-doubling (FD) appearance. Based on the overlapping areas of the stimuli we transformed the C20 results to 9 measures for direct comparison with R9. We also compared mfVEP-based and psychophysical contrast-thresholds in 26 younger (26.6 ± 7.3 y, mean ± SD) and 20 older normal control subjects (66.5 ± 7.3 y) control subjects using the R9 stimuli.

Results The best intraclass correlations between R9/C20 thresholds were for the central and outer regions: 0.82 ± 0.05 (mean ± SD, $p \leq 0.0001$). The areas under receiver operator characteristic plots for C20 and R9 were 1.0 ± 0.0 and 0.99 ± 0.012 (mean ± SE) respectively. Canonical correlation analysis (CCA) showed significant correlation ($r=0.638$, $p=0.029$) with 1 dimension of the C20 and R9 data, indicating that the lower and higher temporal frequency tests probed the same neural mechanism(s). Low signal quality made the contrast-threshold

mfVEPs non-viable. The resulting mfVEP thresholds were limited by noise to artificially high contrasts, which unlike the psychophysical versions, were not correlated with age.

Conclusion The lower temporal frequency R9 stimuli had similar diagnostic power to the FDT-C20 stimuli. CCA indicated the both stimuli drove similar neural mechanisms, possibly suggesting no advantage of FD stimuli for mfVEPs. Given that the contrast-threshold mfVEPs were non-viable we used the present and published results to make recommendations for future mfVEP tests.

Keywords: glaucoma, frequency doubling, perimetry, neural mechanisms, multifocal VEPs

Introduction

The Frequency Doubling Technology (FDT) and Matrix perimeters have been a relatively successful addition to our tools for managing glaucoma, with reports of earlier detection [1], and better correlation with nerve fibre loss [2]. That being said more recent studies suggest that FD based perimetry does not perform very differently to other standard methods [3,4]. The FDT C20 program uses 0.25 cpd gratings that are contrast modulated at 25 Hz to generate the spatial frequency doubling (FD) illusion. FD stimuli were hypothesised to preferentially stimulate the nonlinear Y-like cell population which, in having a very low coverage factor, are a good target for glaucoma diagnosis compared to other ganglion cell populations in which up to 20 cells see each point in visual space [5]. Y-cells have now been identified in primates [6-8] but the Y-cell hypothesis remains contentious [9,10]. Electrophysiology experiments have reported quite good glaucoma discrimination performance using similar low spatial frequency stimuli but modulated at 8 to 10 Hz [11,12] where FD is reported to be seen with lower probability [13]. Here we seek to provide evidence as to whether we should shift our electrophysiological stimuli towards FD conditions.

Rosli et al. [13] mapped the spatio-temporal conditions for seeing frequency doubling at different locations within the visual field. Intermediate spatial frequencies are reported for suboptimal conditions. This has been interpreted as the visual system resolving the concurrent presence of FD and non-FD responses in different add-mixtures [14]. That being said Rosli et al. [13] provided evidence for up to three independent generators of FD responses covering somewhat overlapping spatio-temporal domains. The eight large spatial regions explored in that psychophysical study were close in size and layout stimuli to those

we have used in several multifocal electrophysiology studies [15-18] including studies of glaucoma [15] and multiple sclerosis [19]. An advantage of these radially scaled stimulus arrays is that, due to cortical magnification, their visual evoked potential responses do not change over 3 or more octaves of viewing distance [17], meaning the same stimuli can be used to objectively test macular and peripheral fields. Collectively those studies suggest that it might be worth using fewer larger stimuli for perimetric testing, especially as fewer larger stimuli reduce psychophysical test times commensurately [20]. Indeed even larger stimuli have been used for contrast-threshold testing in glaucoma with relative success and these had very low test-retest variance [21].

The C20 threshold test of the FDT perimeter examines 17 large test stimuli within the central ±20 degrees of the visual field. We therefore decided to compare that test program to a 9-region (R9) stimulus array that is similar to our previous electrophysiology and psychophysical test patterns [15-17,22]. Thus we wanted to see if a potentially quicker 9-region test was as accurate as the FDT, and which might also be implemented in a practical electrophysiological version. To compare the methods we mapped the C20 data to nine equivalents of the R9 thresholds [19] and examined intraclass correlations between the thresholds arising from the two methods. We also tested normal subjects to allow comparison of the diagnostic power of the methods. For added interest we reduced the temporal frequencies used for R9 to the band 7.14 to 10.4 Hz for comparison with previous studies. We also scaled the spatial frequencies with eccentricity to better match retinal magnification. Given these differences we then examined if the R9 and the C20 thresholds contained independent diagnostic information about glaucoma, which could be used to improve diagnostic power [18]. Collectively these results inform future mfVEP methods in

terms of the best stimuli for glaucoma. Finally we examine a particular mfVEP variant using the R9 stimuli that sought to find objective contrast-thresholds via a stair-case procedure.

Methods

Subjects

Normal subjects had corrected vision of 6/6 or better measured using the Bailey-Lovie chart, refractive errors ≤ ± 6 DS and ± 1.50 DC, normal visual fields were assessed by the screening program of the C-20-5 program of the FDT perimeter (Model 710, Carl Zeiss Meditec, Dublin, CA). None of the normal subjects, had a self-reported history of glaucoma, or was under medication for any systemic or ocular disorders.

Optometrists and resident glaucoma sub-specialists (including GFA, MAH) classified the glaucoma patients' eyes using a predefined scale of glaucoma severity to classify them into suspect, mild, moderate and severe categories based upon the HFA 24-2 mean defects (MD), with cut-offs at 6 and 12 dB, and other features [18]. Vertical cup-to-disc ratios (VCDR) were determined by slit-lamp examination and IOP by applanation tonometry. Details of how the HFA, IOP and VCDR data were used by the clinicians to define the diagnostic groups has been published as Table 2 of [18]. All glaucoma was primary open angle glaucoma (POAG). Since we wanted to explore the full gamut of disease this cross-sectional study also included glaucoma suspects with IOP > 21 mm and vertical cup/disc ratios of 0.6 to 0.7. Inclusion criteria for the patients were: best-corrected visual acuity of 6/12 or better, spherical refractive error of 0 ± 6 DS, astigmatism within 0 ± 3 DC, no more than mild cataract.

The exclusion criteria for all subjects were retinal dystrophies, mixed ocular disease, dense cataract, previous intraocular surgery except for uncomplicated glaucoma or cataract

surgery, or any systemic disease or medication likely to affect vision. Eyes with pupil diameter less than 2 mm, were also excluded. Normal subjects with risk factors for glaucoma, or any other eye diseases that may affect the visual field (VF) were excluded.

The eye with the best acuity was tested. In the event of equal acuities an eye was selected at random. For R9 the untested eye was occluded with a black patch, to emulate the C20 method where the untested eye is in darkness. Subjects were not refracted to the viewing distance for either method due to the very modest demodulation of grating stimuli below 1 cpd [23,24].

[Figure 1 about here please]

Main tests and stimuli

In each of its 17 regions (Fig. 1C) the C20 test displayed 0.25 cpd gratings that were contrast reversed at 25 Hz. The stimuli were presented for a maximum of 720 ms, with 160 ms onset and offset ramps, and a variable period at the maximum contrast of 200 to 400 ms. Contrast-thresholds were determined by a modified binary search (MOBS) method based upon 6 reversals.

The nine-region (R9) stimuli were generated by a 24-bit Vista graphics card (Truevision, Shadeland Station, IN) and were displayed on a HP 1230 CRT monitor at a resolution of 512 by 424 pixels measuring 333 mm by 417 mm. Subjects viewed the stimuli from a distance of 370 mm. The monitor display rate was 101.5 frame/s. The sinusoidal contrast modulation frequency for each region differed within a narrow range between 7.14 and 10.4 Hz, where the frequencies of bilaterally equivalent regions had frequencies that differed by < 0.6 Hz (Fig. 1B).

The mean luminance of the R9 screen was 51.8 cd/m^2, similar to C20 at 100 cd/m^2. Contrast-thresholds were based upon 6 reversals of a 4-2 staircase method [25,26]. The range of sinusoidally modulated grating contrasts was fixed to 16 semi-octave steps of either 1.25 or 1.76 dB. Further details of the threshold method are available [18]. The starting contrast was -11.0 dB (i.e. 8%), which has previously been shown to be efficient [27]. The stimulus duration was 1 s. A fixation cross was presented within a 1 degree disc at 0 contrast, and the boundaries of the central and innermost rings of stimuli were at radius 3 and 10 degrees (Fig. 1A,B). The outer boundary was at 24 degrees. The spatial frequencies scaled with eccentricity being 1.0, 0.5 and 0.25 cpd (Fig. 1B). The stimulus regions were separated by 0.5 deg. A previous multifocal-VEP study using very similar stimuli showed that even with no gaps the effects of scatter were minimal [16]. Fig. 1C illustrates the C20 test regions, and Fig. 1D illustrates the overlap of the R9 and C20 stimuli.

Analysis

To compare the thresholds from the two methods we applied area-based weights to the FDT C20 stimuli to map them to nine R9 equivalents. The method has been described elsewhere [19]. Comparisons of the R9 and C20 thresholds used Intraclass Correlation Coefficients (ICC) and Bland-Altman plots [28,29].

The *relative* diagnostic power of the R9 and C20 thresholds was investigated using Receiver Operator Characteristic (ROC) plots of sensitivity on the false positive rate. The plots were summarized using the area under the curve (AUC), and we estimated standard errors in the AUCs [30]. All patients were pooled for that analysis. We used a Leave-One-Out (LOO) method in which the distributions of data for the controls were reformed leaving out each control, so that no control was ever classified using their own data.

A Canonical Correlation Analysis was done to investigate whether the C20 and R9 stimuli generate independent (diagnostically complementary) information about glaucoma. All analysis was done using Matlab (release R2016b; MathWorks Inc, Natick MA).

Contrast-threshold mfVEP demonstration

As part of a commentary on the suitability of the R9 stimuli for mfVEPs we examined a mfVEP method using the same spatial and temporal frequencies as R9. The experiments were conducted on 26 younger (26.6 ± 7.3 y, mean ± SD) and 20 older normal control subjects (66.5 ± 7.3 y). The method used the same staircase threshold scheme but where the decision rule was not a button-press but a decision based on the achieved VEP signal-to-noise ratio. We have published all the details of that method [18] in a study comparing psychophysical and VEP thresholds obtained for a single large grating. Here the nine regions were presented simultaneously modulated at the nine temporal frequencies of Fig. 1B. As we have shown previously this allows the response to the nine regions (and possible interactions) and independent noise amplitudes within the 3.26 Hz band to be extracted from the discrete Fourier transform. Details of those mfVEP methods are given elsewhere [17,22]. Thus, the staircase proceeded in parallel for the 9 regions with each mfVEP recording taking 20.2 seconds. These data were collected on 46 younger and older normal controls to examine the feasibility of an mfVEP based threshold method for R9-like stimuli. Psychophysical contrast-thresholds were also determined for each of these subjects using the same stimuli.

Results

Table 1. Subject demographic data and the number of eyes tested (1 per subject) for the R9/C20 psychophysical experiments. Errors are SD.

	CONTROLS	PATIENTS
NUMBER	16	19
FEMALE	11	7
OD	12	9
AGE	67.8 ± 5.65	71.9 ± 7.15

There were 16 control subjects and 19 patients and their demographic data are summarised in Table 1. The number of males to females, tested right eyes or left, and the ages of the controls and patients were not significantly different. We wished to compare performance over a range of disease severities, therefore 7 patients were POAG suspects, 2 had mild POAG in their tested eye, 5 had moderate POAG, and 5 had severe disease (Methods). No separate analysis by disease category was done.

Figure 1 shows the layout and sizes of the two types of stimuli used, each of which displayed contrast modulated low spatial frequency gratings (Methods). Fig. 1A, B illustrate the R9 stimulus areas and types. Fig. 1C shows the layout of the 17 FDT C20 test stimuli, and Fig. 1D illustrates how the two types of stimuli overlap. Fig. 1D shows that the central stimuli of the two methods overlap considerably. Similarly the outer 9 regions of the C20 stimuli substantially overlap with the outer 4 regions of the R9 pattern. We used weights based upon the overlapping areas of the two types of stimuli to create nine R9-equivalent thresholds from the 17 C20 thresholds/field (Methods).

[Figure 2 about here please]

Fig. 2A to C are scatter plots of the central, middle and outer ring threshold data (see legend) comparing the R9 and mapped C20 data. Fig. 2F summarises the per-region Intraclass Correlation Coefficients (ICCs). Even correcting for multiple comparisons the correlations of the central and outer four regions were significant at $p \leq 0.0001$. The significance of the correlation for the middle four regions ranged from $p = 0.001$ to $p = 0.020$.

Fig. 2A shows that the central R9 regions gave less sensitive thresholds on average than the larger central C20 stimulus by 15.6 ± 3.85 dB (mean ± SE). By contrast outer regions produced 4.54 ± 1.93dB more sensitive thresholds (Fig. 2C). These features are better illustrated by Bland-Altman plots (Fig. 2 D,E) and their associated linear models. B-A plots are more appropriate than scatter plots when comparing methods or test-retest variance, reducing the effect of regression to the mean by plotting the differences (R9-C20) on the means of R9 and C20. The linear models fitted a constant and a slope (green dash-dot lines in 2 D,E) to the B-A data. The constant characterises a parallel shift from the diagonals of Fig. 2A-C and the slope characterises any heteroskadatic behaviour. Summary results are shown in Table 2. Overall there was good agreement between the methods.

Table 2. Bland-Altman plot summary data. Constant and slope characterise the green dash-dot lines in Fig. 2 D,E. For the inner and outer regions the p-values are corrected for multiple (4) comparisons. As expected from Fig. 2C the Centre data showed a significant constant shift. The dimensionless Slope component characterised significant heteroskadastic behaviour for the Centre and Inner regions. The upper and lower 95% confidence limits (-95CL, +95CL) represent the horizontal dashed black lines in Fig. 2 D,E. The SD diff columns is the standard deviation in the R9-C20 differences.

Row	Constant (dB)	p	Slope	p	-95CL	+95CL	SD diff
Centre	-15.6 ± 3.85	0.00	0.30 ± 0.15	0.05	-21.26	4.86	6.66
Inner	5.81 ± 2.56	0.10	-0.25 ± 0.09	0.03	-15.52	13.83	7.49
Outer	4.54 ± 1.93	0.08	-0.01 ± 0.07	1.00	-7.34	15.81	5.91

We next compared the *relative* diagnostic power of the methods by examining the area under Receiver Operator Characteristic plots (AUC of ROC) for discriminating the patients from control subjects. Note that this analysis is not intended to say anything about the ability of either test to diagnose glaucoma, we are only examining the relative ability of the two tests to discriminate the two subject groups. The LOO method also reduced the issue of a small number of controls (Methods). To examine how diagnostic power varied with the number of test-regions considered we recomputed the ROCs nine times (for each method), first using the single worst performing test region in the field of each patient and control, and then for the mean of the worst 2, 3 ,4,…9 regions. Thus all 9 ROCs were based on 1 measurement per eye. Fig. 3 shows the resulting AUCs and their 95% confidence limits. AUC was highest for the single worst point indicating that very few controls had a region that showed very poor sensitivity, while even the glaucoma suspects did. Diagnostic power dropped as more regions were averaged, as would be expected if an increasing number of normal regions were being recruited into the means for each subject. The AUC values for the means of 8 and 9 worst regions from the two methods differed at $p \leq 0.05$, but not when corrected for multiple comparisons.

[Figure 3 about here please]

The R9 and C20 stimuli used somewhat different spatial and temporal frequencies and produced somewhat different results (e.g. Fig. 2). A question therefore is do the two methods contain independent (uncorrelated) information about the severity of glaucomatous change? Such information can be used to improve diagnostic power. As in a previous study [18] we examined this question using Canonical Correlation Analysis (CCA). The ROC analysis suggested that for both methods a reasonable variable would be the

region that had the lowest sensitivity. We therefore formed a matrix, M, of functional data with two columns, comprising the minimum sensitivity from each method for each subject's eye. We then formed a disease severity matrix, G, having columns that were a 4-point severity scale (suspect, mild, moderate, severe), or a binary scale that grouped suspects and mild, and moderate and severe eyes. The approach for G was suggested by our published results [18], which indicated that different tests can carry independent information about early- vs. later-stage disease. We then entered M and G into the CCA.

As a first step the CCA confirmed that each matrix was of full rank (i.e. 2) indicating that each contained two independent forms of information. The CCA then formed two matrices of canonical variates, U and V, which are akin to the principal components of M and G, and computed the correlations between these. The first components of U and V were correlated at r=0.638 (p=0.029), but not the second components (r=0.178, p=0.411). This indicated that there was only one type of information in M that was related to glaucoma using either test method. Predictably the first component of the transformed functional data U (a linear combination of the R9 and C20 minima in M) was significantly correlated with the two original severity variables at r=0.636 and r=0.567 (p=0.001 and p=0.005), but the second component of U was not. Thus, it appeared that, despite the differences in spatial and temporal frequencies, the two test methods did not provide significantly different types of information about glaucoma, but rather they measured much the same information implying that they tested the same neural mechanism, or more than one mechanism in about the same proportions.

Contrast-threshold mfVEP demonstration

[Figure 4 about here please]

We have published a study that used the same staircase threshold method as the R90 experiments above to find contrast-thresholds for a single large grating using either button presses, or VEP signal-to-noise ratios to determine the subjects' "responses" [18]. Here we attempted the same approach for the R9 stimuli (Methods) in a group of 26 younger (26.6 ± 7.3 y, mean ± SD) and 20 older normal control subjects (66.5 ± 7.3 y). Both psychophysical and mfVEP versions were tested in all subjects. The outcome (Fig. 5) was that the smaller VEP signal-to-noise ratios afforded by the R9 stimuli (compared to a single large grating of the former study) meant that the mfVEP method tended to settle on an artificially high "threshold" level of about 5 dB (30% contrast). Thus, while the psychophysical R9 method of this paper showed much lower thresholds of 2 to 3% contrast (15 to 17 dB) that were significantly different between the two age groups ($p<0004$), the R9-like mfVEP threshold data was not different between age groups. There was also no significant correlation between the psychophysical and mfVEP data for either age group.

Discussion

Overall the C20 and R9 psychophysical methods gave very similar thresholds (Fig. 2) and had similar diagnostic power (Fig. 3). The CCA indicated that both methods were essentially measuring the same neural system or systems. The lower thresholds for the central region of the R9 stimuli (Fig. 2A) might suggest that a central stimulus with somewhat higher spatial frequency than used by C20 might be preferable, as retinal and cortical magnification would suggest [17]. That psychophysical data and the poor performance of the contrast-threshold mfVEP, taken together with published results on similar mfVEPs, can collectively inform us about the best stimuli and methods for future mfVEP studies.

Frequency doubling

[Figure 5 about here please]

In a previous study we used very similar sized regions to R9 to explore the spatial and temporal domain over which the FD percept is seen [13]. In 17 subjects we tested 35 spatial/temporal frequency combinations in each of 8 visual field regions, which were similar to the 8 outer regions of the R9 pattern. The conclusion of that study was that there were up to 3 independent neural mechanisms subserving perception of FD. An approximation of the overlapping spatio-temporal ranges of those mechanisms is shown in Fig. 5, where the (tested) spatial frequencies on the abscissa correspond to the inner regions of R9, and half those values for the outer regions. Even at low contrasts FD was reported down to 9.36 Hz across all the spatial frequencies (cf. Figures 4 and 6 of [13]). This would suggest that the thresholds reported here might test at least two of those mechanisms (M1 and M3 of Fig. 5). CCA indicated only a single significant type of information about glaucoma severity, but this could due to similar stimulation of M1 and M3 by both tests. We used a

similar CCA analysis of thresholds derived from psychophysical and VEP methods applied to the same subjects [18]. In that study the stimulus was a single large 1 cpd grating modulated at 7.2 Hz. Those two threshold types carried information on two independent sources of diagnostically useful information about glaucoma, and combining them increased AUC levels.

Relevance to mfVEPs

The CCA analysis here indicated no great value for glaucoma diagnosis for FD stimuli as opposed to stimuli using lower temporal frequencies: apparently similar mechanisms are tested by both. A caveat for VEPs is that as temporal frequency is increased to around 20 Hz the SNR improves due to decreasing noise (e.g. Fig 3a of [19]). We also concluded that the contrast-threshold R9 mfVEP method was not worth pursuing. Instead any future mfVEP studies with R9-like stimuli should use the more conventional method of measuring the average response to several repeats at a single indicative contrast (e.g. [15,16,19]), even 10 repeats taking well under 4 minutes to potentially measure both eyes [16]. Two of our previous mfVEP studies, using stimuli that were very like R9, showed that contrast-response functions tend to plateaux by 30 to 60% contrast [17,22]. Thus to avoid the effects of saturation the constant test contrast should be at about these levels. Just two repeats of the 20.2 second stimuli at 48% contrast yielded a median of 4 out of 9 regions significant at $p<0.05$ (Fig. 6A of [17]), even when using higher temporal frequencies that gave somewhat smaller responses than those used here. Doubling the contrast made little difference.

Small vs. large stimuli

One advantage of using fewer larger stimuli than *standard automated perimetry* (SAP) is test duration can be reduced [20]. Another advantage is that large stimuli provide lower

test-retest variability than SAP (reviewed [31,32]). The very high test-retest variability of SAP is a problem for tracking disease progression. We have provided evidence that the Goldmann Size 3 stimuli of SAP may produce much of the retest variance due to the interaction between normal fixational jitter and spatially rapid changes in the visual fields of glaucoma patients [31,32]. Another reasonable suggestion for a source of retest variance is larger time-varying noise in unhealthy retinal ganglion cells [33]. A recent study isolating the time-varying and spatial components of retest variance indicated however that the spatial component is about 12 times larger [34].

Partially as a way of overcoming these difficulties normal clinical practice is to employ averages of SAP thresholds as used by the glaucoma hemifield index (GHT). The GHT pools data from between 3 and 6 test points of the 24-2 test grid to generate 5 scores per hemifield. This common averaging strategy has been demonstrated by key glaucoma opinion leaders to be an effective way of interpreting SAP data more generally [35].

Of course very large stimuli, or averages of SAP stimuli like the GHT, could reduce the detection rate of small scotomas. By contrast 24-2 or 30-2 programs of SAP test less than 0.5% of the area within each 6 degree cell of those test arrays, thus potentially missing small- to medium-sized scotomas altogether. Perhaps the best example of this is a study comparing the ability of the HFA 30-1 and 30-2 test patterns (off-set by 3 degrees) to detect scotomas in 100 glaucomatous eyes [36]. The study found that one or the other pattern often missed or underreported significant scotomas, leading to poor correlation between the two methods.

Conclusions

Some potential limitations of this study were the relatively small number of subjects, the somewhat different temporal frequencies/R9 region, and the different threshold strategies for R9 and C20. Overall a test strategy like R9 was shown to be reasonable and future study is probably merited. The study allows specific recommendations for future mfVEP stimuli and methods to be made.

Figure legends

Fig. 1 Stimulus layout and details. **A** The layout of the R9 stimulus array where the grey levels assist to visualise the different regions. **B** The spatial and temporal frequencies of the R9 stimuli. From centre to periphery the spatial frequencies were 1, 0.5 and 0.25 cpd (any FD frequencies would be 2, 1, and 0.5 cpd). The different temporal frequencies were chosen to allow a future multifocal VEP experiment, and were very similar in bilaterally symmetric parts of the field. **C** The layout of the 17 stimuli of the C20 program, the thin brighter vertical bars correspond to where the C20 stimuli overlap (if they were ever shown simultaneously but they are not). **D** Shows the overlap of the C20 and R9 regions. It is simply the sum of the grey levels of A and C with an adjustment to make the regions all relatively visible.

Fig. 2 Agreement of the thresholds for all subjects for the 9 different parts of the visual field. The units for the axes of A to E are decibels. **A** The central visual field thresholds show significant correlation but the C20 stimuli illicit overall larger thresholds (text). **B** Scatter plot of the threshold from the inner 4 regions. **C** Scatter plot of the results from the outer four regions. Note that for A to C the translucent dots allow overlapping data points to be seen and the additive colour of the overlapping dots is indicative of density. **D** Bland-Altman plot for the inner regions. The green dashed-dot line is a fit to a linear model, the parameters of which are given in Table 2 along with the data for the central and outer regions. **E** Bland-Altman plot for the outer regions. **F** Summary of the C20/R9 intraclass correlation coefficients (ICCs) for each region.

Fig. 3 Summary of 9 ROC analyses for each of R9 and the C20 mapped data showing the mean ± 95% confidence limits for the Area under each ROC curve (AUC). For the N-worst = 1 analysis the region of the field with the lowest sensitivity (worst) was selected, distributions for the control subjects and patients were formed, and the ROC curve constructed. N-worst = 2 indicated that the 2 worst regions of each field were found and averaged. N-worst = 3 to 9 indicate means across the worst 3 to 9 regions/field, and so the ROCs for all nine cases were based that single number per eye. Thus N-worst = 9 indicates the overall mean threshold for each field. A Leave-One-Out method was used in the ROC analysis so that no control subject was classified using a control distribution that included their own data. After correction for multiple comparisons none of the points in the two curves was significantly different from its fellow point.

Fig. 4 The contrast-thresholds achieved for the R9 stimuli using both the psychophysical method (Psych), and an mfVEP method made possible by the different temporal frequencies of R9 (Fig. 1B). Both methods were tested on younger and older (abscissa labels = mean ± SD) normal controls. The psychophysical thresholds were quite sensitive and differed between the groups but the mfVEP thresholds were dominated by noise, providing no information about age, and were not correlated with psychophysical thresholds.

Fig. 5 Approximate spatiotemporal domains of independent neural mechanisms (M1 to M3) that operate where the FD illusion is seen for each region of a very similar stimulus array to the R9 pattern [13]. In fact they overlap somewhat. The data suggest that the R9 and C20 threshold data might contain add-mixtures information from M1 and M3. If the add-

mixtures were different (and especially if nonlinearities were involved) then the two types of threshold might contain independent information about glaucoma that would increase diagnostic power (e.g. ROC AUC). CCA indicated that either: C20 and R9 were subserved by a single mechanism, or more than one mechanism in similar add-mixtures.

Compliance with ethical standards

Conflict of interest

Some readers will know that Maddess held the patents for the FDT/Matrix perimeters, however those patents lapsed in 2015 and Maddess retains no interest in those Carl Zeiss products. Indeed the R9 method of the paper might be seen as a competitor of FDT. Maddess has a small holding in, and is on the advisory board of, EyeCo Pty Ltd, which is developing treatments for retina oedema and dry macular degeneration. He could also earn royalty income from patents assigned to Konan Medical USA Inc for a pupillography based perimetry system, but which has no features like R9 or FDT. R9 could be seen as a competitor to that potential product. Author SNA declares that she has no conflict of interest. Author GFS is deceased. Author MAH declares that he has no conflict of interest.

Funding

This research was supported by the Australian Research Council through the ARC Centre of Excellence in Vision Science (CE0561903) and intramural funding from the Australian National University.

Ethical approval and Informed Consent

All procedures performed in studies involving human participants were in accordance with the ethical standards of the University of New South Wales and the University of Otago; and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

References

1. Medeiros FA, Sample PA, Weinreb RN (2004) Frequency doubling technology perimetry abnormalities as predictors of glaucomatous visual field loss. Am J Ophthalmol 137:865-871

2. Kim TW, Zangwill LM, Bowd CB, Sample PA, Shah N, Weinreb RN (2007) Retinal nerve fiber layer damage as assessed by optical coherence tomography in eyes with a visual field defect detected by frequency doubling ecthnology perimetry but not by standard automated perimetry. Ophthalmology 114:1053–1057

3. Liu S, Lam S, Weinreb RN, Ye C, Cheung CY, Lai G, Lam DS, Leung CK (2011) Comparison of standard automated perimetry, frequency-doubling technology perimetry, and short-wavelength automated perimetry for detection of glaucoma. Invest Ophthalmol Vis Sci 52 (10):7325-7331. doi:10.1167/iovs.11-7795

4. Tafreshi A, Sample PA, Liebmann JM, Girkin CA, Zangwill LM, Weinreb RN, Lalezary M, Racette L (2009) Visual function-specific perimetry to identify glaucomatous visual loss using three different definitions of visual field abnormality. Invest Ophthalmol Vis Sci 50 (3):1234-1240. doi:10.1167/iovs.08-2535

5. Maddess T, Henry GH (1992) Performance of nonlinear visual units in ocular hypertension and glaucoma. Clinical Vis Sci 7 (5):371-383

6. Petrusca D, Grivich MI, Sher A, Field GD, Gauthier JL, Greschner M, Shlens J, Chichilnisky EJ, Litke AM (2007) Identification and characterization of a Y-like primate retinal ganglion cell type. J Neurosci 27 (41):11019-11027. doi:10.1523/JNEUROSCI.2836-07.2007

7. Crook JD, Peterson BB, Packer OS, Robinson FR, Gamlin PD, Troy JB, Dacey DM (2008) The smooth monostratified ganglion cell: evidence for spatial diversity in the Y-cell pathway to the lateral geniculate nucleus and superior colliculus in the macaque monkey. J Neurosci 28 (48):12654-12671. doi:10.1523/JNEUROSCI.2986-08.2008

8. Crook JD, Peterson BB, Packer OS, Robinson FR, Troy JB, Dacey DM (2008) Y-cell receptive field and collicular projection of parasol ganglion cells in macaque monkey retina. J Neurosci 28 (44):11277-11291. doi:10.1523/JNEUROSCI.2982-08.2008

9. White AJ, Sun H, Swanson WH, Lee BB (2002) An examination of physiological mechanisms underlying the frequency-doubling illusion. Invest Ophthalmol Vis Sci 43 (11):3590-3599

10. Swanson WH, Sun H, Lee BB, Cao D (2011) Responses of primate retinal ganglion cells to perimetric stimuli. Invest Ophthalmol Vis Sci 52 (2):764-771. doi:10.1167/iovs.10-6158

11. Marx MS, Podos SM, Bodis-Wollner I, Lee PY, Wang RF, Severin C (1988) Signs of early damage in glaucomatous monkey eyes: low spatial frequency losses in the pattern ERG and VEP. Experimental Eye Research 46 (2):173-184

12. Johnson MA, Drum BA, Quigley HA, Sanchez RM, Dunkelberger GR (1989) Pattern-evoked potentials and optic nerve fiber loss in monocular laser-induced glaucoma. Invest Ophthalmol Vis Sci 30 (5):897-907

13. Rosli Y, Bedford SM, Maddess T (2009) Low spatial frequency channels and the spatial frequency doubling illusion. Invest Ophthal Vis Sci 50:1956-1963. doi:doi:10.1167/iovs.08-1810

14. Kulikowski JJ, Maddess T Apparent finess of compound gratings. In: Invest. Ophthal. Vis. Sci., (Ft. Lauderdale, USA), 1998. p 405

15. Maddess T, James AC, Goldberg I, Wine S, Dobinson J (2000) Comparing a parallel PERG, automated perimetry and frequency doubling thresholds. Invest Ophthalmol Vis Sci 41:3827-3832

16. Rosli Y, Maddess T, Dawel A, James AC (2009) Multifocal frequency-doubling pattern visual evoked responses to dichoptic stimulation. Clinical Neurophys 120:2100-2108

17. Abdullah SN, Aldahlawi N, Vaegan, Boon MY, Maddess T (2012) Effect of contrast, region number and viewing distance on multifocal steady-state visual evoked potentials (MSVs). Invest Ophthalmol Vis Sci 53:5527-5535

18. Abdullah SN, Sanderson G, James AC, Vaegan, Maddess T (2014) Visual evoked potential and psychophysical contrast thresholds in glaucoma. Doc Ophthalmol 128:111-120

19. Ruseckaite R, Maddess T, Danta G, James AC (2006) Frequency doubling illusion VEPs and automated perimetry in multiple sclerosis. Doc Ophthalmol 113:29-41

20. Johnson CA, Samuels S (1997) Screening for glaucomatous visual field loss with frequency doubling perimetry. Invest Ophthalmol Vis Sci 38:413-425

21. Maddess T, Goldberg I, Dobinson J, Wine S, Welsh AH, James AC (1999) Testing for glaucoma with the spatial frequency doubling illusion. Vision Res 39:4258-4273

22. Abdullah SN, Vaegan, Boon MY, Maddess T (2012) Contrast-response functions of the multifocal steady-state VEP (MSV) Clin Neurophys 123:1865-1871

23. Campbell FW, Green DG (1965) Optical and retinal factors affecting visual resolution. J Physiol 181 (3):576-593

24. Horner DG, Dul MW, Swanson WH, Liu T, Tran I (2013) Blur-resistant perimetric stimuli. Optom Vis Sci 90 (5):466-474. doi:10.1097/OPX.0b013e31828fc91d

25. Henson DB (2000) Strategies used in examining the visual field. In: Visual Fields. Reed Educational and Professional Publishing Ltd, Oxford, pp 23-41

26. Cubbidge R (2005) Threshold strategies. In: Doshi S, Harvey W (eds) Eyes Essential, Visual Fields. First edn. Elsevier Butterworth-Heinemann, pp 24-35

27. Vaegan, Rahman AMA, Sanderson GF (2008) Glaucoma affects steady state VEP contrast thresholds before psychophysics. Optom Vis Sci 85 (7):547-558

28. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1 (8476):307-310

29. Swanson WH, Malinovsky VE, Dul MW, Malik R, Torbit JK, Sutton BM, Horner DG (2014) Contrast sensitivity perimetry and clinical measures of glaucomatous damage. Optom Vis Sci 91 (11):1302-1311. doi:10.1097/OPX.0000000000000395

30. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143 (1):29-36. doi:10.1148/radiology.143.1.7063747

31. Maddess T (2011) The influence of sampling errors on test-retest variability in perimetry. Invest Ophthalmol Vis Sci 52:1014-1022

32. Maddess T (2014) Modelling the relative influence of fixation and sampling errors on test-retest-variability in perimetry. Graefes Arch Ophthalmol 252:1611–1619

33. Gardiner SK, Swanson WH, Goren D, Mansberger SL, Demirel S (2014) Assessment of the reliability of standard automated perimetry in regions of glaucomatous damage. Ophthalmology 121 (7):1359-1369. doi:10.1016/j.ophtha.2014.01.020

34. Numata T, Maddess T, Matsumoto C, Okuyama S, Hashimoto S, Nomoto H, Shinomura Y (2017) Exploring test-retest variability using high-resolution perimetry. Trans Vis Sci Tech 6 (5(8)):1-9. doi:10.1167/tvst.6.5.8

35. Johnson CA, Sample PA, Cioffi GA, Liebmann JR, Weinreb RN (2002) Structure and function evaluation (SAFE): I. criteria for glaucomatous visual field loss using standard automated perimetry (SAP) and short wavelength automated perimetry (SWAP). Am J Ophthalmol 134 (2):177-185

36. Weber J, Dobek K (1986) What is the most suitable grid for computer perimetry in glaucoma patients? Ophthalmologica 192 (2):88-96

Figure 1

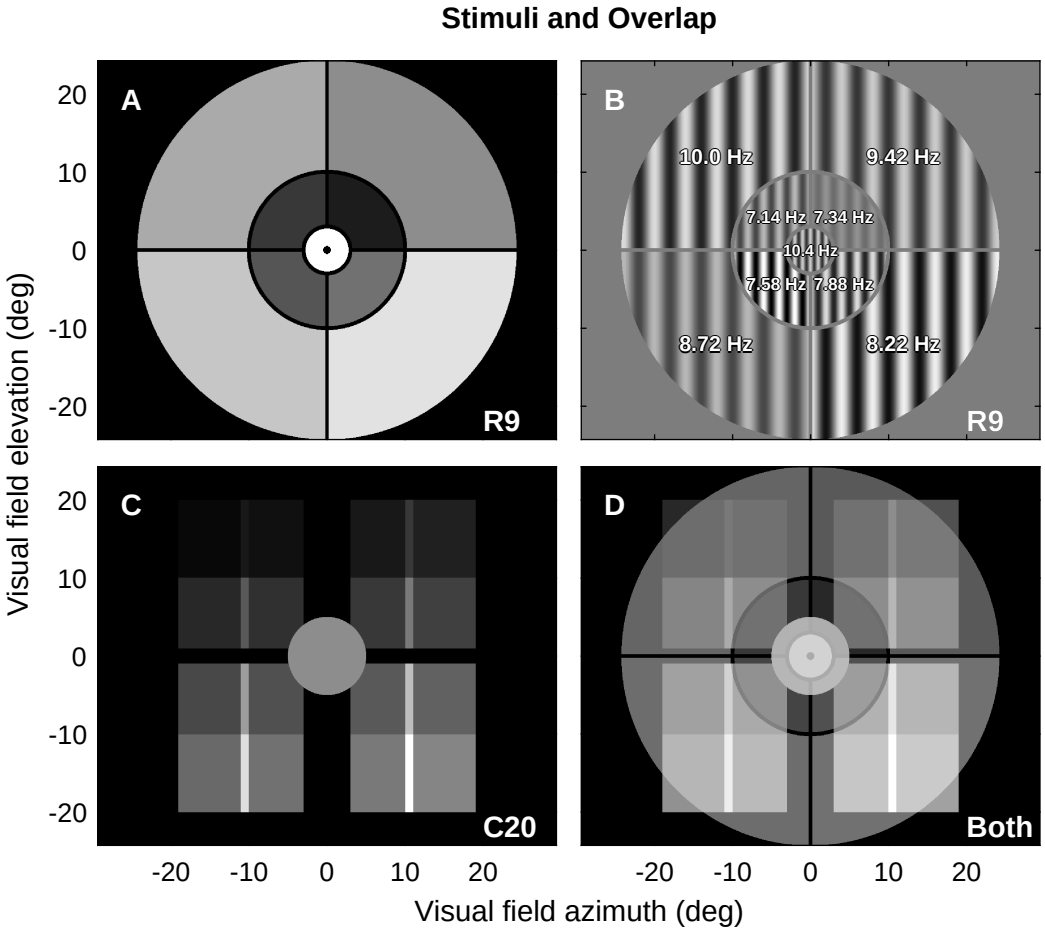
Stimuli and Overlap
A
R9
B
10.0 Hz
9.42 Hz
7.14 Hz
7.34 Hz
10.4 Hz
7.58 Hz
7.88 Hz
8.72 Hz
8.22 Hz
R9
C
C20
D
Both
Visual field elevation (deg)
20
10
0
-10
-20
-20
-10
0
10
20
Visual field azimuth (deg)

Figure 2

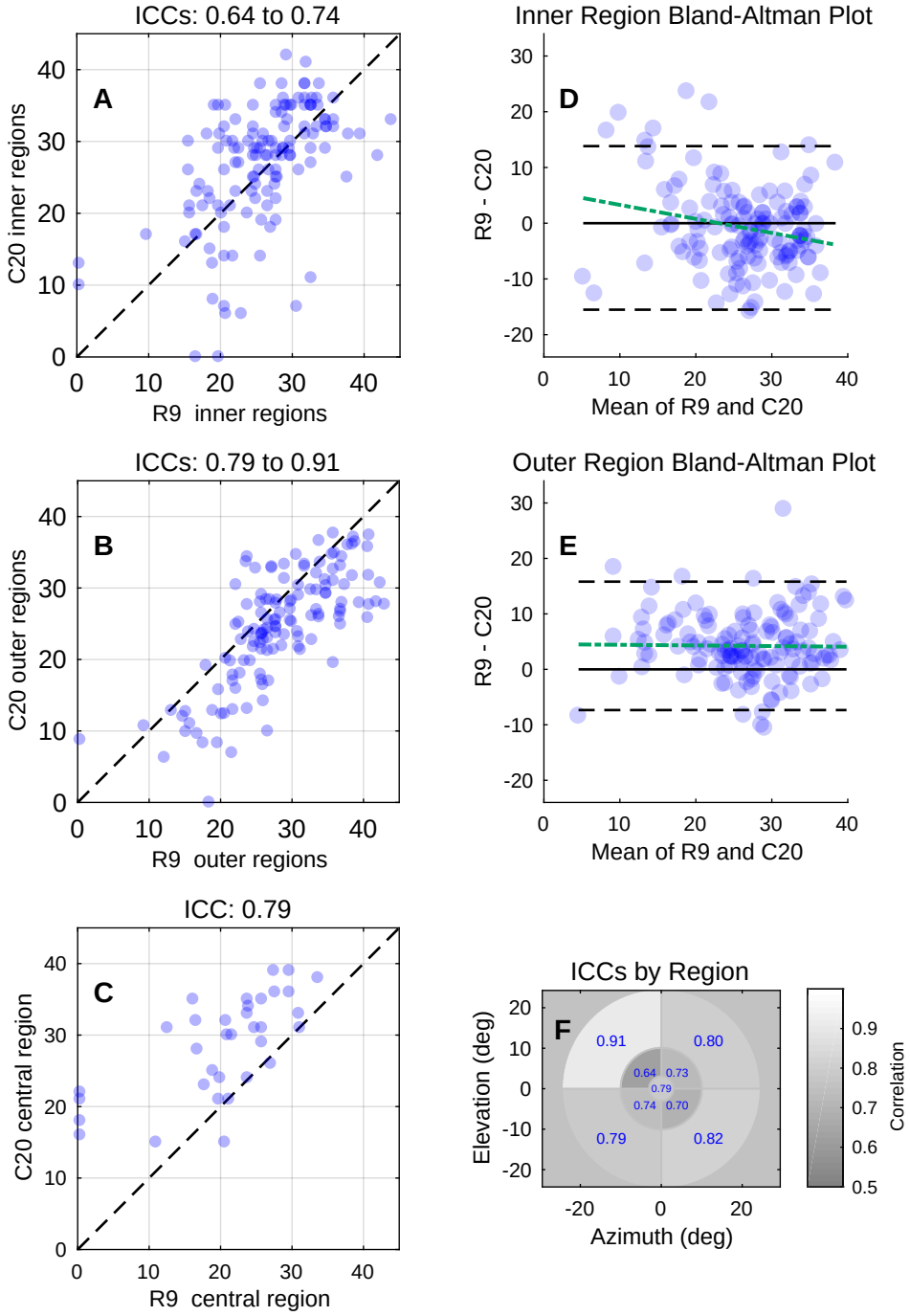
ICCs: 0.64 to 0.74
A
C20 inner regions
R9 inner regions
Inner Region Bland-Altman Plot
D
R9 - C20
Mean of R9 and C20
ICCs: 0.79 to 0.91
B
C20 outer regions
R9 outer regions
Outer Region Bland-Altman Plot
E
R9 - C20
Mean of R9 and C20
ICC: 0.79
C
C20 central region
R9 central region
ICCs by Region
F
0.91
0.80
0.64
0.73
0.79
0.74
0.70
0.79
0.82
Elevation (deg)
Azimuth (deg)
Correlation

Figure 3

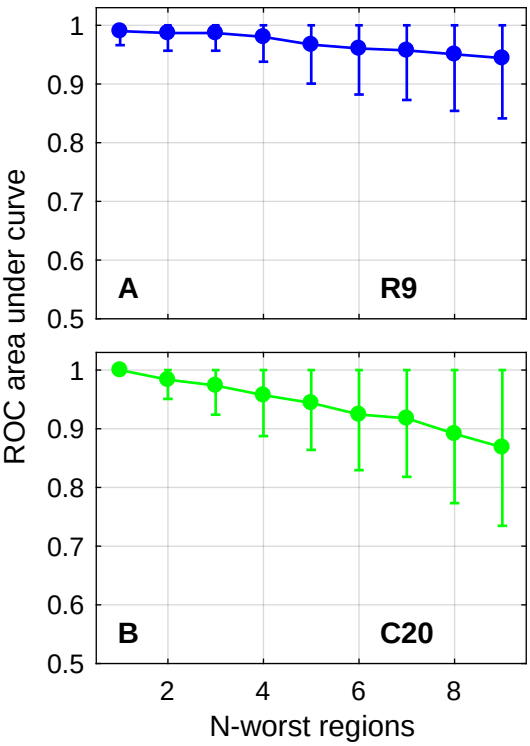
ROC area under curve
1
0.9
0.8
0.7
0.6
0.5
A
R9
1
0.9
0.8
0.7
0.6
0.5
B
C20
2
4
6
8
N-worst regions

Figure 4

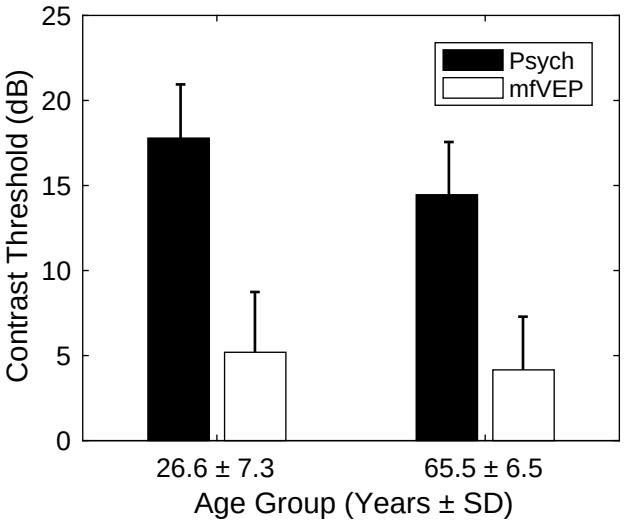
25
20
15
10
5
0
Contrast Threshold (dB)
Psych
mfVEP
26.6 ± 7.3
65.5 ± 6.5
Age Group (Years ± SD)

Figure 5

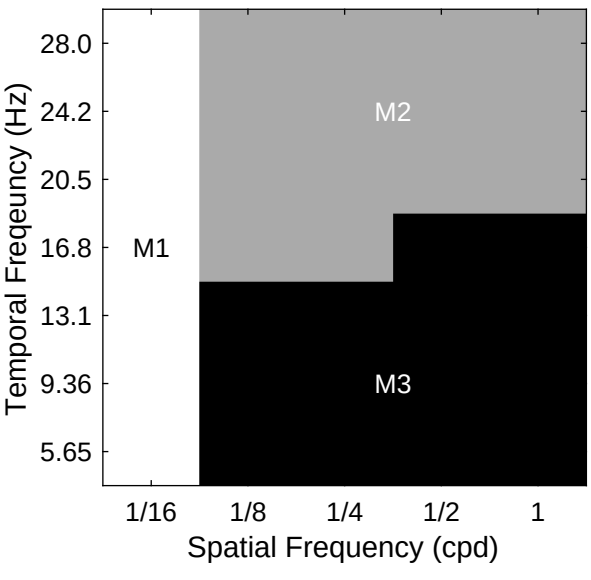
28.0
24.2
20.5
16.8
13.1
9.36
5.65
Temporal Freqeuncy (Hz)
M1
M2
M3
1/16
1/8
1/4
1/2
1
Spatial Frequency (cpd)

Table 1. Subject demographic data and the number of eyes tested (1 per subject) for the R9/C20 psychophysical experiments. Errors are SD.

	CONTROLS	PATIENTS
NUMBER	16	19
FEMALE	11	7
OD	12	9
AGE	67.8 ± 5.65	71.9 ± 7.15

Table 2

Table 2. Bland-Altman plot summary data. Constant and slope characterise the green dash-dot lines in Fig. 2 D,E. For the inner and outer regions the p-values are corrected for multiple (4) comparisons. As expected from Fig. 2C the Centre data showed a significant constant shift. The dimensionless Slope component characterised significant heteroskadastic behaviour for the Centre and Inner regions. The upper and lower 95% confidence limits (-95CL, +95CL) represent the horizontal dashed black lines in Fig. 2 D,E. The SD diff columns is the standard deviation in the R9-C20 differences.

Row	Constant (dB)	p	Slope	p	-95CL	+95CL	SD diff
Centre	-15.6 ± 3.85	0.00	0.30 ± 0.15	0.05	-21.26	4.86	6.66
Inner	5.81 ± 2.56	0.10	-0.25 ± 0.09	0.03	-15.52	13.83	7.49
Outer	4.54 ± 1.93	0.08	-0.01 ± 0.07	1.00	-7.34	15.81	5.91