

Original Article

The influence of sampling errors on test-retest variability in perimetry

T Maddess

ARC Centre of Excellence in Vision Science, Centre for Visual Sciences, Research School of Biology, Australian National University, Canberra, Australia.

Disclosure: T. Maddess, Carl Zeiss Meditec (P); Seeing Machines (F,C,P)

Corresponding author:

Prof. Ted Maddess
ACEVS and Centre for Visual Sciences
The Australian National University
Canberra ACT 0200
Australia
ted.maddess@anu.edu.au

Running Head: sampling and test-retest

Description: The changes in SAP test-retest variability reported for different scotoma depths and stimulus sizes are replicated by considering the effects of under-sampling and normal variations in fixation.

Keywords: automated perimetry, test-retest variability, aliasing, under-sampling

Word count: 3805 words

Abstract

Purpose. To determine if visual fields measured by standard automated perimetry (SAP) can be distorted by higher spatial frequency image components; in particular if test-retest variability of SAP fields can be explained by the combination of small scale fixational drift, small stimulus size, and coarse spatial sampling of the visual fields.

Methods. Standard SAP test patterns have points 6 degrees apart. The amplitude spectra of 511 Humphrey Field Analyser (HFA) 10-2 fields were assessed to see if their finer grained sampling revealed spatial frequencies which could cause distortions of standard fields due to under-sampling. Model visual fields were then constructed whose spectra were similar to the 10-2 fields and test-retest variability was examined for Goldmann Size III to VI stimuli and Gaussian fixational drift with standard deviations of 0.075 to 0.3 deg.

Results. The 10-2 fields showed significant spatial frequency content up to 0.25 cpd, three times the highest frequency that a 30-2 or 24-2 sample grid can resolve. As reported for SAP, test-retest variability increased with scotoma depth, and increasing the stimulus size from Size III to VI caused a reduction in test-retest variability, as did reduced fixation jitter.

Conclusions. Using fixation drift half the size of that exhibited by good fixators many of the features of SAP test-retest variability were reproduced. Reducing test-retest variability may therefore involve using large test stimuli that are blurry in appearance, and which overlap somewhat when placed on the perimetric test grid. Overlap across the meridians should perhaps be avoided.

247 words

Introduction

Test-retest variability is a major problem for all forms of standard automated perimetry (SAP) including the achromatic¹⁻³ and blue/yellow programs⁴⁻⁵ of the Humphrey Field Analyser (HFA), the Octopus perimeter,⁶ and also the FDT perimeters.⁷⁻⁸ It is often supposed that much this variability arises from issues of patient vigilance and attention. Some evidence for this is that 5 to 10% of subjects fail quality controls such as high rates of false positive responses and fixation losses.⁹⁻¹³ However, some investigations indicate patient issues are not the main sources of variability in perimetry.¹⁴ High test-retest variability is problematic because it interferes with our ability to track visual field progression. Thus for example it is recommend that new patients be given 6 field tests over the first two years to have a 80% chance of detecting mean loss of -2 dB/yr.¹⁵ Is it possible that non-patient related issues also affect the reliability of visual fields?

There is some evidence that stimulus size is a factor. The original frequency doubling stimuli were very large and the standard deviation for test-retest was about 1 dB in patients and normal subjects.¹⁶ The first FDT perimeter had stimuli 10 deg square and had lower test-retest variability than SAP,⁷ however halving the stimulus size for the 24-2 pattern of FDT2 increased variability⁸. On a similar note Wall et al.² [IOVS 2009;50: E-Abstract 2239] have shown that for achromatic SAP variability decreases with increasing stimulus size, from Goldmann Size III to Size VI.

Part of the problem associated with small stimuli may be related to the very coarse grid of SAP test points, which are spaced at 6 deg intervals for standard HFA 24-2 and 30-2 test patterns. Thus the standard Size III stimuli test probes less than 0.41% of the test pattern

sampling and test-retest

area ($0.146/6^2 * 100$). This means much can be missed. The problem may be much worse, however, if the sensitivity variations across a damaged visual field are not very smooth.

A grid of sample points spaced at 6 deg intervals can only accurately reconstruct spatial frequencies up to the Nyquist frequency:¹⁷ $1/12$ (0.083) cpd, or about 4 cycles across a 24-2 field. If higher spatial frequencies than 0.083 cpd exist in the field, and the sampling grid is quite orderly, then the higher frequencies will manifest themselves as lower frequencies in the band 0 to 0.083, those image components being said to be *aliased*.¹⁷⁻¹⁸ Moire patterns are examples of the capacity of such aliased image components to distort a sampled image. Random sampling arrays produce related problems.¹⁹ If aliased components distort the measured fields then small variations in fixation would create new distortions on each retest. Micro-perimetry studies indicate that *good* fixators have deviations from fixation described by a distribution with a standard deviation (SD) of 0.6 degrees.²⁰⁻²¹

This study first examines 10-2 fields, with their 3 times higher sampling density, to ask whether significant image components exist above the Nyquist rate for 24-2 and 30-2 fields. Such components could form distorting aliases within measured visual fields. The study then examines the effect on the measured test-retest variability of eye movements by modelling tiny deviations from the test grid between repeated field tests that mimic normal eye movements during fixation. The outcomes indicate that most of the results reported for SAP test-retest variability, including the reported changes with stimulus size, can largely be predicted by aliasing effects.

Methods

All data analysis and modelling was performed using Matlab (The Mathworks, Natick, MA).

The HFA 10-2 visual field data was de-identified data collected under the Australian National University's Human Experimentation Ethics Committee protocol 04/238, and conformed to the declaration of Helsinki.

Fourier analysis

Visual field data from right eyes was flipped left to right to make them equivalent to left eye data to permit a common analysis. The threshold data from each field was then inserted into a matrix of zeros that was 33 by 33 pixels in size. The amplitude spectrum of each field, Af_i , was then computed. That is to say the 2D-Fourier transform of each of the fields was computed and the absolute value taken, the resulting spectrum characterizes the amplitude (in dB) and orientation of every spatial frequency in the visual field map. It is worth mentioning that any image, including a visual field map, can be exactly represented as a sum of sinusoidal gratings, hence the spectrum identifies what spatial frequencies are found in the image from which it was derived.¹⁷ As will be described in the text these spectra were then averaged by *groups*, Af_g , defined by features such as the presence or absence of large nasal steps. Next amplitude spectra of flat fields with the same mean sensitivity as each group, Ff_g , were computed. The differences $Df_g = Af_g - Ff_g$ were then taken to find the components of the Af_g that were not due to the windowing the data caused by setting the points outside the field to 0. Samples, along lines through the various Df_g , so called transects, were then taken to examine Fourier image components that were oriented horizontally, vertically or along the 45 degree diagonal of the fields. So for example vertical meridian of

sampling and test-retest

the 2D spectrum describes the amplitudes (strengths) of all the spatial frequencies contained within the visual field whose spatial modulation is oriented vertically.

Model fields

Part of the study required model visual field data that were generated on a finer spatial scale than a conventional SAP test pattern. Depending on the particular study the spatial resolution of these model fields, Δs , was 1/10 or 1/30 degree. The requirement was for a pattern with diffuse concentrations of damage mimicking scotoma as well as more normal parts of the field that were very smooth. A further constraint was the amplitude spectra of these model visual fields should be similar to those observed for the real 10-2 visual field data. To do this a matrix of resolution Δs was first initialized with random numbers with a uniform distribution and then a large median filter, ranging from 1.5 deg to 4.5 deg on a side, was operated one on the random data. The resulting data were then scaled from 0 to 1 and then the square root was taken. The resulting model visual fields, m_i , were then scaled between about 4 and -35 dB to imitate the range of SAP thresholds.

Sampling

Sampling employed disc shaped operators, o_i , the size of Goldmann Size III, V or VI stimuli. Note that for a viewing distance of 300 mm the radii of these stimuli in degrees are $[0.216 \ 0.862 \ 1.724] = \text{atan}(\sqrt{z^2/\pi}, 300) * 180 / \pi$, where $z=[2 \ 6 \ 8]$. The volume of the operator was set to 1. Sampling then involved locating the operator over the model fields, m_i , multiplying the points in the operator with those underlying them in m_i and taking the sum. Since the volume of $o_i = 1$ the final sum was the mean of the threshold data under the operator. For any given sampling of 24-2 points the exact position of o_i was jittered by a circular Gaussian distance with SD equal to values ranging between 0.075 and 0.3 degrees depending on the

study. All the jitter sizes used were thus 2 to 8 times smaller than the 0.6 deg SD reported for good fixators.²⁰⁻²¹ Strictly, because the decibel values were averaged this corresponds to computing the geometric mean of the underlying linear visual field sensitivities. For completeness we also did studies in which the decibel thresholds were first transformed to linear sensitivities, the sample means within the stimuli were computed, and then the results back-transformed to decibel measures. Thus in these cases the arithmetic mean of the sensitivity, rather than the geometric mean, was computed. As expected either method yielded very similar results.

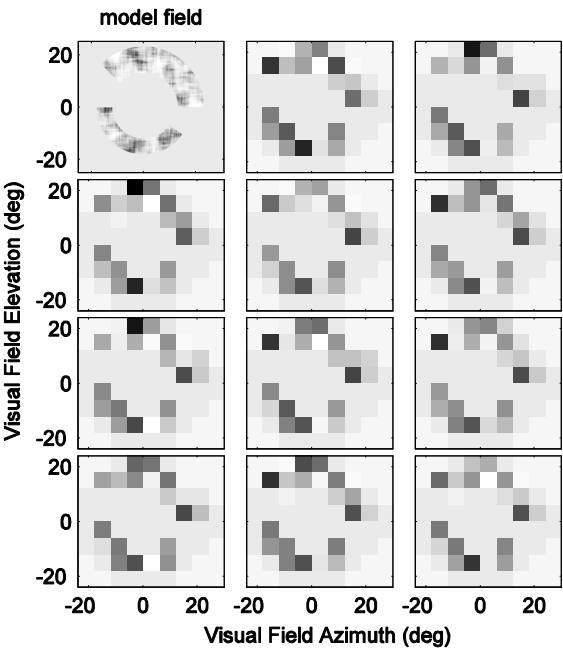


Figure 1. Top left: a model visual field showing two arcuate scotomas containing regions of patchy damage. Horizontal transects through the upper and lower scotomas are shown in Fig. 2A. The remaining figures are the output of sampling the model field with 24x2 grid pattern plus positional errors described by a circularly symmetric Gaussian function of standard deviation 0.3 deg. The median filter used to create the field was 3.1 deg on a side. It is noticeable that very dark and light parts of the sampled fields modulate quite markedly, in much the way that repeated visual field tests do. The gray scale for the sampled fields spans -20 to 2 dB, for the model field -28 to 4 dB, the difference in scale arising from the smoothing effects of the Size III stimulus.

Results

The potential problem is illustrated in Figure 1. At top left is a model visual field created as described in Methods. The field sensitivity data was windowed to have two arcuate scotomas. The other panels are the result of sampling the model field using a Goldmann Size III stimulus at the standard test positions of a 24-2 field, but where those positions have been jittered to mimic small deviations from fixation. The jitter was described by a 2D Gaussian distribution with a standard deviation of 0.3 deg. That is to say the exact positions at which the field was sampled were displaced from the true 24-2 grid positions by distances described by a small circularly symmetric Gaussian probability function. Thus 63.4% of the displacements were within 0.3 deg (1 SD) and 98.2% within 0.6 deg (2 SD) from the nominal 24-2 positions.

Figure 2A shows two horizontal transects through the upper and lower scotomas of the model field in order to give an impression of the sensitivity fluctuations in the model. The test-retest variability from 100 samplings from the same model field is shown in Figure 2B. The variability increases for more damaged parts of the field, much as has been reported for repeated SAP fields.^{1-2, 8}

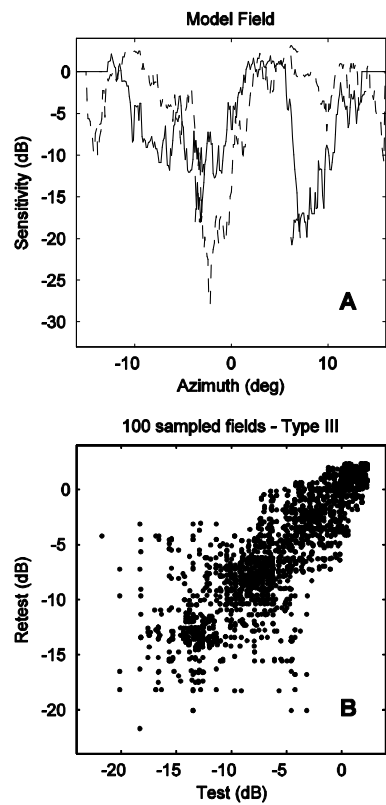


Figure 2. A) two horizontal transects through the model field at top left of Fig. 1. The solid line is taken from the upper scotoma and the dashed line through the lower scotoma. B) Test-retest measurements from 100 samples from the same field using a Size III stimulus with a fixation jitter of 0.3 deg (SD). Each point represents a test-retest pair from some part of the field. Variability grows with scotoma depth as is commonly reported for SAP fields using a Size III stimulus.

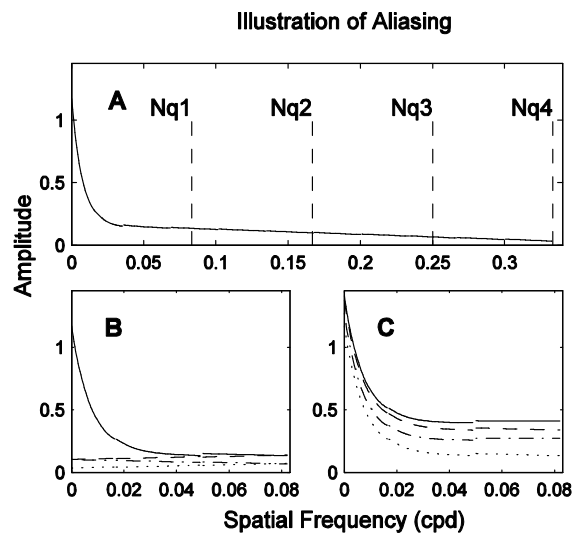


Figure 3. A) shows a putative amplitude spectrum of an image, such as a visual field sensitivity map, which is to be sampled by a 6 deg grid of points. The frequency axis extends to $4Nq_{24-2}$, slightly more than the Nyquist frequency for a 10-2 field with its sampling period of 2 deg (i.e. $Nq_{10-2} = 0.25$ cpd). Multiples of Nq_{24-2} are marked by vertical dashed lines labelled Nq1 to Nq4. The aliasing effect seen within the band 0 to Nq_{24-2} is that the spatial frequency content of the image above Nq_{24-2} ($=Nq1$ in A), is folded back into the range below Nq_{24-2} . B) shows the folding back and forth of the spectrum, thus the higher frequencies masquerade as lower frequencies and are referred to as *aliased frequencies*. (C) The folded back frequencies add, the cumulative additions being shown as the different lines with the final sum being the black line. Thus in practice a considerable part of an under-sampled image can be made up of jumbled, erroneous, projections of the true image.

The situation is perhaps worse than might be expected when considered in the context of sampling a 2D object with a substantially regular grid of sample points. For the 24-2 grid the separation is 6 deg. This defines a Nyquist sampling frequency, Nq_{24-2} , of $1/(2 \times 6) = 0.083$ cpd. That is to say the highest spatial frequency that a 6 degree sampling array can accurately report on is a sinusoidal variation with a period of 12 degrees, or about 4 cycles across a 24-2 field. For the square 24-2 sampling grid Nq is even lower along the 45 and 135 deg diagonal axes, being $1/(2 \times \sqrt{2 \times 6^2}) = 0.059$ cpd, a spatial frequency with period 17.0 deg. From the perspective of perimetry, even if the sample grid is not jittered, this poses a problem if the sampled field contains variations in sensitivity that modulate faster than 4 cycles per visual field diameter. Given the punctuate nature of glaucomatous visual field

defects more rapid variations than 0.083 cpd across the field seem likely. If such image variations exist then the problem of *aliasing* arises, this is where spatial frequencies higher than Nq masquerade as lower frequencies. The aliasing process can be understood as a folding back of the spectrum of the image about Nq_{24-2} , is illustrated in Figure 3.¹⁷⁻¹⁸ The situation is somewhat more complex for slightly jittered sampling arrays but the principle is the same, the sampled image will contain erroneous modulations creating false impressions of the true visual field. This will be exacerbated if, on repeated visits, the sampling grid varies even by a small amount (Fig. 2A).

The question therefore arises: “do real visual fields contain any higher spatial frequencies than Nq_{24-2} that might present a problem as illustrated in Figures 1 and 2A?”. A related issue is what would a field look like if it contained no information above Nq_{24-2} ? A few papers report visual field data sampled at one degree intervals for either whole fields,²²⁻²³ or small parts of fields.²⁴⁻²⁵ A beautiful example from Sturmer²² is shown in Figure 4B where an inferior arcuate scotoma is resolved into several islands of damage. Panel C shows the same data filtered to remove all content above 0.1 cpd (slightly above Nq_{24-2}) indicating what a field would need to look like if 6-degree sampling was dense enough to prevent aliasing. Figure 4B is much more in keeping with our expectations of the structure of a glaucomatous visual field than is Figure 4C.

Figure 4A shows data from the amplitude spectrum of 4B. The solid line is the mean of the 6 vertical transects on either side of the vertical midline of the 2D spectrum (error bars are 95% confidence limits). The dash-dot line is a similar average of the 6 transects bracketing the horizontal midline of the spectrum. It is clear that there is significant information up to at least 2.5 times the Nyquist rate. We also examined data from four 10 degree square fields

sampled at 1 degree intervals from Westcott et al.²⁴. All those visual fields samples contained significant content above Nq_{24-2} .

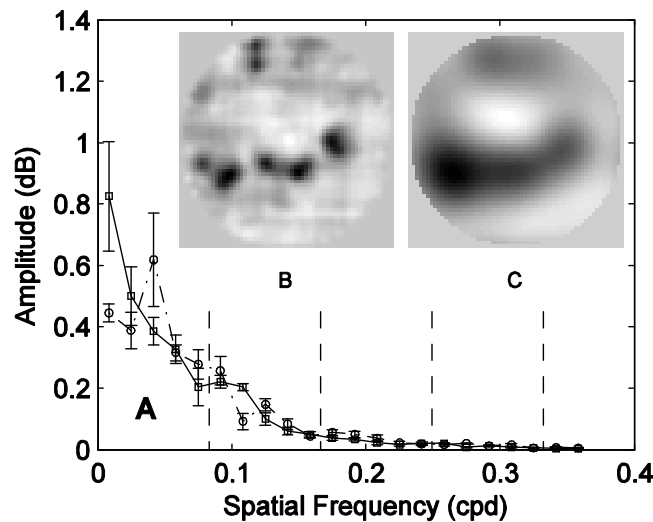


Figure 4. A) The means of 6 transects on either side of the vertical (solid) and horizontal (dash-dot) meridians of the 2D amplitude spectrum of the visual field of B. B) A visual field sampled using a square grid with separation 1 deg. C) The data of B filtered to have no content above 0.1 cpd. If 6-degree sampling were adequate then glaucomatous fields would need to look like this figure. In practice the more rapidly varying picture of B is more in keeping with our understanding of the structure of visual fields. The vertical dashed lines show multiples of Nq_{24-2} as in Fig. 3A.

To address this issue more quantitatively 511 HFA 10-2 fields were examined. For these fields Nq is 3 times higher given the 2 deg separation of the 10-2 sample points. Figure 5 presents transects through the mean 2D spectra of different selections of the 511 fields. The three transects were the horizontal and vertical meridians of the spectra, and also a transect along the 45 deg diagonal of the spectra (see legend in Fig. 5A). Figure 5A shows the average of these transects and SE for all 511 2D amplitude spectra. The spectrum of a normal flat field was subtracted (Methods), thus the spectra in Figure 5 represent deviations from a normal flat field. The Nq_{24-2} for the horizontal, vertical or diagonal sampling of the field by a 24-2 array (0.083 and 0.059 cpd) are shown as dashed vertical lines below 0 dB. Above the Nq_{24-2} values the amplitudes are typically less than 1 dB but are many SE away from 0 and so

are highly significant. The larger values for the vertical transect (the dash-dot line) are undoubtedly due to the fact that many of these visual fields had steps.

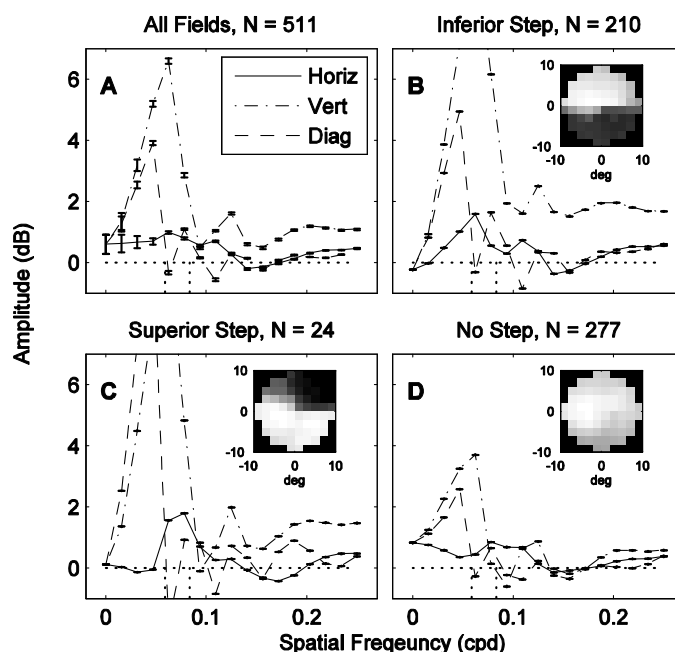


Figure 5. A) transects through the average 2D amplitude spectra for 511 HFA 10-2 fields. The legend indicates the transects that are the horizontal (Horiz) and vertical (Vert) meridians of the 2D spectra, and also transects taken from the 45 deg diagonal of the spectra (Diag). The Nq_{24-2} values for square and diagonal sampling (0.083 and 0.059 cpd) are indicated as vertical dotted lines below 0 dB. Transects for separate averages are shown for fields with steps (B,C), and the 277 fields without steps (D). The insets are the means of each type of field. For all types of field the horizontal meridian transect is unaffected by modulations of the field in the vertical direction such as nasal steps. The diagonal transects may be somewhat affected by such steps depending on their exact shape. The SE are shown for every point, and in many cases are so small that they cannot be distinguished from the transect lines. It is clear that many of the spectra have amplitudes around 0.5 to 1 dB above Nq_{24-2} , which are many SE away from 0 dB and which are not due to nasal steps.

For this reason fields with (Fig. 5B,C) and without steps (Fig. 5D) also had their spectra averaged separately. While the vertical meridians of the spectra are likely to be affected by any features like nasal steps, this is not true for the horizontal meridian transects. The transects along the diagonal of the spectra may be partially affected by step-like features depending on the exact shapes of the defects in given fields. It is clear that most of the 12

sampling and test-retest

transects have frequency components whose amplitudes are many SE away from 0 dB above Nq_{24-2} . Therefore it seems that these image components are real and so will be aliased back into visual fields sampled with a 24-2 or similarly coarse array of test points.

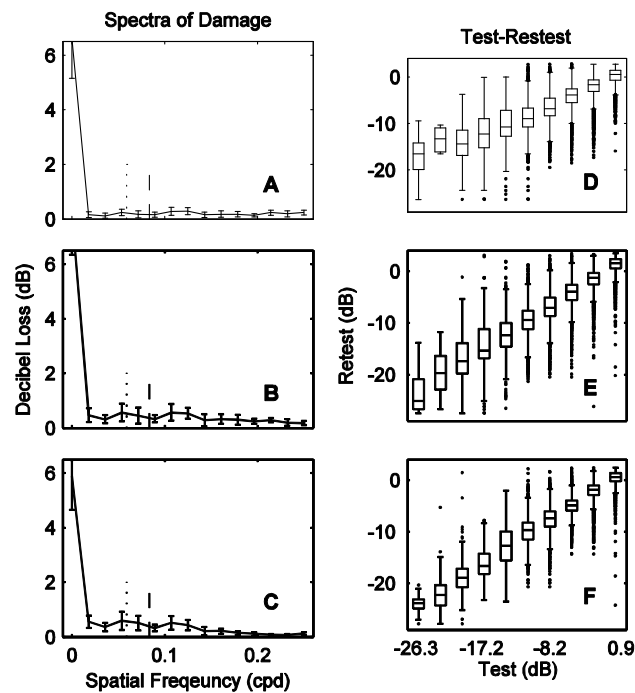


Figure 6. Six model fields were created for each row of the figure. A to C are horizontal transects through the 2D amplitude spectra showing their content out to 0.25 cpd (cf. Figs. 3 and 5). The error bars are standard deviations, and for $N=6$ are essentially 95% confidence limits. For each row the model fields were created using a median filter that was 1.5 deg square (A,D), 3.1 deg square (B,E) and 4.5 deg square (C,F). The test-retest plots (D to F) are each based on 1800 samplings of the model fields, and the small dots beyond the 5th and 95th percentile whiskers are outliers. The modelled fixation jitter was 0.3 deg (SD). The larger median filter sizes increased the low frequency content and reduced the high frequency content. The square and diagonal sampling Nq for 24-2 fields are indicated in A to C by the vertical dotted and dashed lines at 0.083 and 0.059 cpd.

The results also indicated that the spectra of the model fields used here should also be examined. As mentioned in the Methods the model fields were constructed by operating a median filter on random noise and then stretching the results nonlinearly to produce somewhat tighter, deeper, model lesions. Figure 6 illustrates the effect of varying the size of

the median filter on the spectra (Fig. 6 A to C) and also on the corresponding test-retest variability (Fig. 6 D to F), all measured for a Size III stimulus. The test-retest data are shown as box-plots, where the upper and lower margins of the boxes indicate the 25th and 75th percentiles, and the whiskers the 5th and 95th percentiles. This mode of plotting was chosen for comparison with similar plots in the literature characterising test-test variability from 24-2 SAP fields.^{1-2, 8} From the upper to the lower rows of Figure 6 the size of the median filters ranged from 1.5 deg, 3.1 deg and then 4.5 deg on a side. Note that while the median filter will tend to preserve any large scale features that emerge it also hugely reduces the amplitude of rapid variations in the original noise image. Although the different filters altered the frequency content appreciably, the test-retest variability (D to F) only changed a little; the main effect possibly being a slight reduction in variability for deeper defects (cf. Fig. 6E and F). That is consistent with there being less high frequency information in Figure 6F to cause an aliasing for the fields created with the largest filter (cf. Fig. 6B and C). Overall it appears that the assumptions of the model are met providing there is some image content with amplitudes around 1 dB beyond the Nyquist rate for 24-2 fields.

If aliasing were occurring then it should be ameliorated by using larger stimuli because this will effectively blur the sampled image, reducing the high frequency content. Wall et al. have recently a reduction in test-retest variability for Size V and Size VI Goldman stimuli.², [IOVS 2009;50: E-Abstract 2239]. Figure 7 shows the effect of varying the stimulus size on the test-retest variability for model fields. As in Figure 6 two types of model fields were examined, those with median filters of 3.1 (Fig. 7 A to C) and 4.5 deg on a side (Fig. 7 D to F). As seen in Figures 6B,C the larger filter size creates fields with somewhat more low frequency content, consistent with slightly larger aggregations of loss. For both filter sizes the results are basically in accord with the observations of Wall et al. One difference is that

for Size VI stimuli the model fields created with the smaller median filter seem to show some loss of dynamic range (Fig. 7C), whereas those for the larger median filter do not (Fig. 7F). This may indicate that the model fields of Figure 7 D to F are more like those of real fields, and the smaller features in the fields of the 3.1 deg model fields are highly blurred by the 3.45 deg wide Size VI stimuli. As in the data of Wall et al. a floor effect, where no measurement could be less than -30 dB, was introduced here to make the comparison more equivalent.

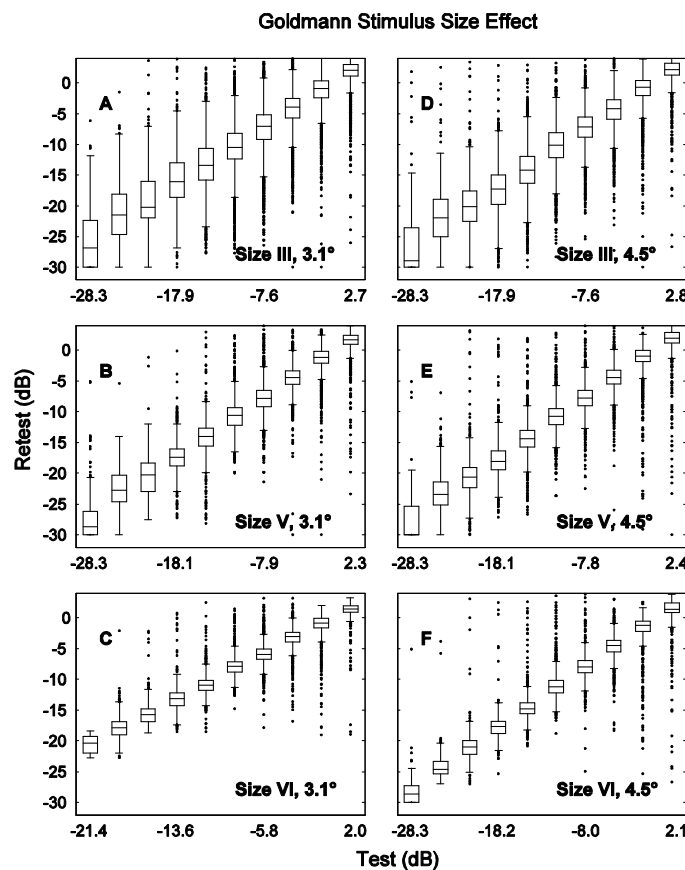


Figure 7. The effect of varying the stimulus size on the test-retest variability measured from model fields. There were 9 model fields per panel and 100 samplings of each field, so that each plot is based on 900 sampled fields. As reported in the literature for SAP fields the variability is reduced with increasing stimulus size from III to V and VI (0.43, 1.72 & 3.45 deg diameter). A to C) the median filter used to create the model fields was 3.1 deg on a side defining a defect size range within the model field. D to F) the median filter size was 4.5 deg indicating larger defects than in A to C. For Size VI stimuli there appears to be a compression of dynamic range in C, but not in F, so it would appear that the large Size VI stimulus is greatly blurring many of the features of the field for fields based on the smaller filter size. Panel F is more consistent with Wall et al. [IOVS 2009;50: E-Abstract 2239], so perhaps real fields are more like the 4.5 deg case. Box plots are as in Fig. 6.

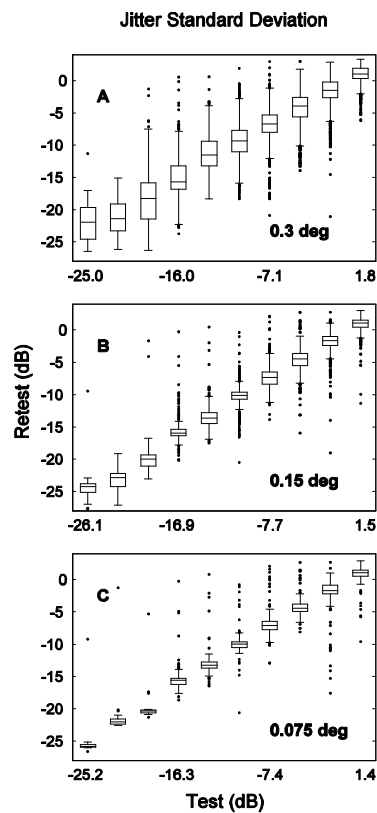


Figure 8. The effect of reducing the standard deviation of the fixation sampling error upon the test-retest variability of model fields. Test-retest variability reduces without a change in dynamic range as jitter is reduced.

A final question is what is the effect of varying the size of the jitter in the sampled fields? To address this question the spatial grain of the model fields was reduced to 1/30 deg, while the median filter size was maintained at 3.1 deg square. The standard deviation of the distribution describing the sampling position error was then reduced from the standard 0.3 deg to, 0.15 deg and then 0.075 deg. Note That the 0.3 deg case (Fig. 8A) is effectively a repeat of Figures 2A, 5E, and 6A computed at higher resolution. Although different model fields and spatial resolutions are used in those figures the outcomes are all very similar.

Discussion

The possibility that test-retest variability seen in many forms of perimetry¹⁻⁸ could be due to undersampling was suggested by Figures 1 to 3. Fourier analysis of data from earlier studies where a few glaucomatous fields were sampled at 1 degree intervals^{22, 24} supported the idea that real fields contain significant spatial frequency content that would form distorting spatial aliases if sampled with a 6 degree grid (e.g. Fig. 4A). Figure 4B,C illustrated the important concept that visual fields smooth enough to not suffer from aliasing effects would look quite unlike our concept of a glaucomatous field. The result that real visual fields contain image components that could form distorting aliases was confirmed by a study of 511 10-2 fields (Fig. 5). Model fields were then constructed where variation between smooth normal parts of the field and areas with concentrations of damage were created using median filters of different sizes. The resulting amplitude spectra were similar to those of real 10-2 fields (cf. Fig. 5 and 6 A to C) and test-retest variability of the model fields was similar to that reported in the literature.^{1-2, 8} The best match to published results and the 10-2 spectra was obtained for median filters that were between 3.1 and 4.5 degrees a side, which create aggregations of damage at about those scales (Fig. 6E,F). Changing the stimulus size from Goldmann Size III to VI produced reductions in test-retest variability consistent with studies of Wall et al.² [IOVS 2009;50: E-Abstract 2239] The match was especially good for the model fields created with the larger 4.5 deg filter, in agreement with Figure 4B and other studies of the finer structure of visual fields.²⁴ The modelling studies of test-retest variability (Figs. 1, 2B, 6, 7) used a fixation error described by a circular distribution with an SD of 0.3 deg, half the known fixation jitter of good fixators.²⁰⁻²¹ So like the spatial frequency content of the model fields the fixation error modelled was

conservative. As predicted, when the fixation error SD was decreased to 0.15 deg, and then 0.075 deg, test-retest variability decreased dramatically (Fig. 8).

The present study could be improved in several ways. One improvement would be to study a collection of very fine grained visual fields. These would probably have to be repeated several times within a few weeks and averaged. This would provide an accurate but slightly blurred set of fields and some data on test-retest variability. Methods like Markov Random Fields, using some fine grained fields as input, could generate a large set of model fields that might be more realistic than those used here.²⁶ That being said the 511 10-2 fields used here already have 3 times higher sampling density than conventional fields, so it is not clear how much would be added. Another approach would be to track the eye movements of persons during perimetric testing, and to examine the correlation between test-retest variability the accuracy of fixation. Tracking accuracy less than 0.1 deg would be needed, and to do that while providing a 60 deg field of view would be technically difficult. Also whether anyone could achieve the required <0.1 deg fixation accuracy is moot (Fig. 8). One could try to control the stimulus position to cancel some proportion of the fixation error, but again that would be technically demanding. It is worth noting such fixation error cancelation would not remove the distorting effects of aliasing, only produce more similar distortions of the field on each visit.

Much of the data on SAP test-retest variability has been collected from glaucoma patients, so a possible conclusion might be that the observed variability is a feature of glaucoma. However investigations comparing visual field variability in glaucoma and optic neuritis by Henson et al.¹⁴ indicate that visual field variability is similar whenever defects are produced by ganglion cell loss. The variability measured by those authors was within-test variability

sampling and test-retest

characterised by fitting frequency of seeing (FOS) curves with a cumulative normal distribution where the standard deviation, σ , is indicative of the variability.^{7, 27-28} Interestingly Henson et al.¹⁴ found that stimulus eccentricity, fixation loss rate, false-positive rate and patient age were not significant determinants of the observed variability, consistent with the some factor like the physical process of sampling, rather than patient vigilance, being the determining influence. Their results were taken to support a hypothesis put forward by other investigators^{7, 28} that the variability was in part due to individual ganglion cells giving variable responses. Aliasing would mimic that effect.

Aliasing is probably not the whole story however, for example in a study using very large FD stimuli the FOS curves are still somewhat broader in patients than in normals,²⁹ although σ was only 1.3 times larger in patients. This is unlike SAP studies where σ is reported to increase by 2 to 3 times.^{7, 14, 27-28} Other factors effecting the signal to noise ratios of perimetric methods are also likely to be involved.³⁰

Based on results using the slightly denser sampling grid of the Competer perimeter Heijl³¹ suggested that few aggradations of damage would be missed altogether by a coarse sampling grid since they tend to be several degrees wide, suggestive of the results of Fig. 6. It is also the case that Airaksinen and Heijl²⁵ have reported arcuate scotomas 1 to 2 deg wide and these would generate large amplitudes at spatial frequencies around 1 cpd, about 10 times the Nyquist rate of a 6-degree test grid. Also a study of combined 30-1 and 30-2 perimetry indicated that 43% of 68 eyes with measureable defects had moderate to severe disagreement between the 30-1 and 30-2 tests.³² Of course some of that disagreement would be due to test-retest variability, but the authors presented evidence that it was often due the 6 degree grid missing defects. As suggested by Heijl³¹ high variability means that

retesting is probably more valuable than finer grained sampling. Recent research verifies this idea.¹⁵

If sampling errors and the Size III stimulus are the problem then the solution to the test-retest problem is larger stimuli not finer grained sampling, or more frequent testing. The issues are how large and are there any other relevant properties required? Recent results from Wall et al. suggest that Size V stimuli actually increase the effective dynamic range of SAP.³³ That being said stimuli with sharply defined features, e.g. edges, are problematic for several reasons.

The first is that refraction needs to be excellent for such stimuli. By comparison stimuli that have a blurred appearance, i.e. containing no spatial frequencies above a few cycles per degree, suffer from limited reduction in physical contrast due to several dioptres of defocus.³⁴ This was a concept behind the original FDT perimeter stimuli, which has been shown to be quite tolerant of mis-refraction.³⁵ The main effect of mis-refraction upon FDT stimuli may be to blur the sharp edges of its stimuli since physics would say that the gratings themselves should be little demodulated. The smaller 24-2 stimuli of the FDT 2 (Matrix) perimeter have a greater content of edge relative to the stimulus area, perhaps biasing patients' attention more to the edge contrast than the grating contrast. This might explain the larger test-retest variability of the newer test compared to the original.⁷⁻⁸

Another problem with sharp edged perimetry stimuli relates back to issues of aliasing. Stimuli with sharp features are defined by high spatial frequencies. Therefore sampling a field with them means that high frequency content of the field tends to be preserved, potentially forming the basis of aliased distortion of the measured visual field map. This is again an argument for using large blurry stimuli.

sampling and test-retest

Fundamentally the aliasing effect arises because the act of sampling produces copies in the spectrum of the sampled object. These copies repeat at intervals of twice the Nyquist rate.^{17-18, 36} If the sampled object is not filtered to have frequencies below the Nyquist rate then the copies overlap in the spectrum, creating the appearance of folded back aliased frequencies (Fig. 3). The solution is to prefilter the sampled object. In perimetry terms this would mean that the test stimuli would need to be composed of sufficiently low spatial frequencies, i.e. blurry, so that the resulting sampled spectra do not overlap. In practice this means blurry stimuli that are so large that they overlap somewhat when presented at the perimetric test grid locations. Large overlap might be unacceptable, however, particularly near the horizontal and vertical meridians of the visual field. A compromise is illustrated in Figure 9. Figure 9 shows a 30-2 like stimulus ensemble using large stimuli that have blurred edges (Fig. 9*). The size of the stimuli means that the blurring skirts of the stimuli overlap somewhat except at the meridians.

Conclusions

Overall many of the parameters of the test-retest variability of SAP visual fields seem to be able to be predicted from sampling errors. The study suggests these errors are exacerbated by the existence of higher spatial frequencies in visual fields than can be reconstructed by the sampling array combined with the use of and the small, sharp-edged, Size III stimuli. One solution would be to use larger smooth-edged stimuli that overlap somewhat except at the central meridians.

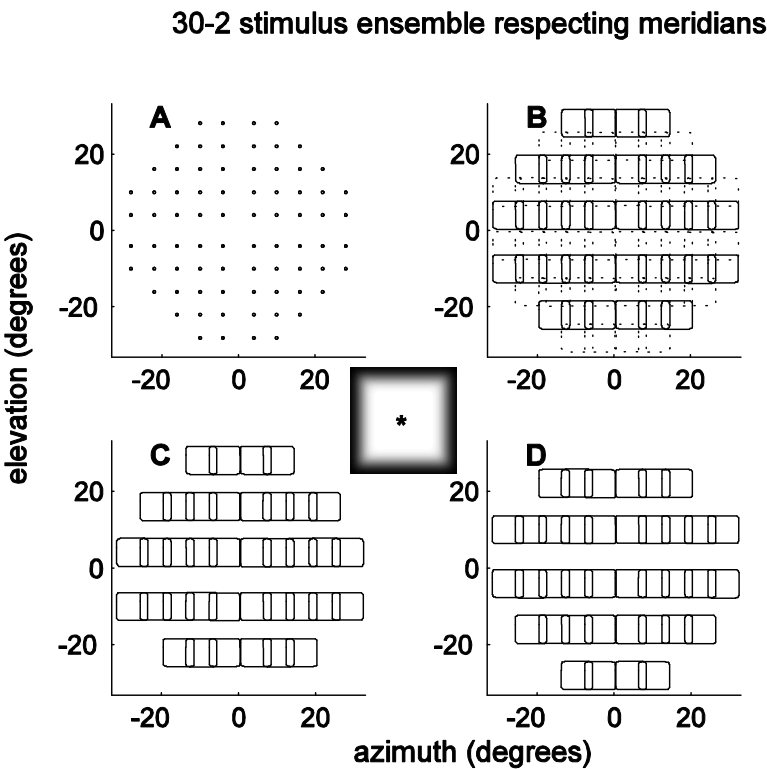


Figure 9. An example of a possible stimulus ensemble that would solve the aliasing problem and provide stimuli that were quite resistant to mis-refraction. *) Illustration of a 9-degree square blurred stimulus, which is suitable for tiling a SAP stimulus grid with minimal overlap. A) the centres of the modified 30-2 grid which are offset by 1 degree from the horizontal and vertical meridians. B) the ensemble of stimuli of the type shown in * centred on the positions shown in A. The size of each stimulus is illustrated by its contour at half height. C,D) every other row of contour plots from B plotted separately to highlight the way the contours respect the meridians.

Acknowledgements

This research was supported by the Australian Research Council through the ARC Centre of Excellence in Vision Science (CE0561903), and the ANU Centre for Visual Sciences. The author is also grateful for insightful comments of the reviewers.

References

1. Artes P, Iwase A, Ohno Y, Kitazawa Y, Chauhan B. Properties of Perimetric Threshold Estimates from Full Threshold, SITA Standard, and SITA Fast Strategies. *Invest Ophth Vis Sci* 2002;43:2654-2659.
2. Wall M, Woodward KR, Doyle CK, Artes PH. Repeatability of automated perimetry: a comparison between standard automated perimetry with stimulus size III and V, matrix, and motion perimetry. *Invest Ophthalmol Vis Sci* 2009;50:974-979.
3. Heijl A, Lindgren A, Lindgren G. Test-retest variability in glaucomatous visual fields. *Am J Ophthalmol* 1989;108:130-135.
4. Blumenthal EZ, Sample PA, Zangwill L, Lee AC, Kono Y, Weinreb RN. Comparison of long-term variability for standard and short-wavelength automated perimetry in stable glaucoma patients. *Am J Ophthalmol* 2000;129:309-313.
5. Kwon YH, Park HJ, Jap A, Ugurlu S, Caprioli J. Test-retest variability of blue-on-yellow perimetry is greater than white-on-white perimetry in normal subjects. *Am J Ophthalmol* 1998;126:29-36.
6. Piltz JR, Starita RJ. Test-retest variability in glaucomatous visual fields. 1990;109:109-111.
7. Chauhan BC, Johnson CA. Test-retest variability of frequency-doubling perimetry and conventional perimetry in glaucoma patients and normal subjects. *Invest Ophthalmol Vis Sci* 1999;40:648-656.
8. Artes PH, Hutchison DM, Nicolela MT, LeBlanc RP, Chauhan BC. Threshold and variability properties of matrix frequency-doubling technology and standard automated perimetry in glaucoma. *Invest Ophthalmol Vis Sci* 2005;46:2451-2457.
9. Bickler-Bluth M, Trick GL, Kolker AE, Cooper DG. Assessing the utility of reliability indices for automated visual fields. Testing ocular hypertensives. *Ophthalmology* 1989;96:616-619.
10. Birt CM, Shin DH, Samudrala V, Hughes BA, Kim C, Lee D. Analysis of reliability indices from Humphrey visual field tests in an urban glaucoma population. *Ophthalmology* 1997;104:1126-1130.
11. Ivers RQ, Macaskill P, Cumming RG, Mitchell P. Sensitivity and Specificity of Tests to Detect Eye Disease in an Older Population. *Ophthalmology* 2001;108:968-975.
12. Pierre-Filho Pde T, Schimiti RB, de Vasconcellos JP, Costa VP. Sensitivity and specificity of frequency-doubling technology, tendency-oriented perimetry, SITA Standard and SITA Fast perimetry in perimetrically inexperienced individuals. *Acta Ophthalmol Scand* 2006;84:345-350.
13. Keltner JL, Johnson CA, Cello KE, et al. Visual Field Quality Control in the Ocular Hypertension Treatment Study (OHTS). *J Glaucoma* 2007;In Press.
14. Henson DB, Chaudry S, Artes PH, Faragher EB, Ansons A. Response variability in the visual field: comparison of optic neuritis, glaucoma, ocular hypertension, and normal eyes. *Invest Ophthalmol Vis Sci* 2000;41:417-421.
15. Chauhan BC, Garway-Heath DF, Goni FJ, et al. Practical recommendations for measuring rates of visual field change in glaucoma. *Br J Ophthalmol* 2008;92:569-573.
16. Maddess T, Goldberg I, Wine S, Dobinson J, Welsh AH, James AC. Testing for glaucoma with the spatial frequency doubling illusion. *Vision Res* 1999;39:4258-4273.
17. Bracewell RN. Chapter 10. Sampling and series., *The Fourier transform and its applications*. New York: McGraw-Hill; 1986:189-218.
18. Giesler WS, Hamilton DB. Sampling-theory analysis of spatial vision. *J Opt Soc Am* 1986;A3:62-70.

19. Maddess T, Hemmi JM, James AC. Evidence for spatial aliasing effects in the Y-like cells of the magnocellular visual pathway. *Vision Res* 1998;38:1843-1859.
20. Rohrschneider K, Becker M, Schumacher N, Fendrich T, Volcker HE. Normal values for fundus perimetry with the scanning laser ophthalmoscope. *Am J Ophthalmol* 1998;126:52-58.
21. Springer C, Bultmann S, Volcker HE, Rohrschneider K. Fundus perimetry with the Micro Perimeter 1 in normal individuals: comparison with conventional threshold perimetry. *Ophthalmology* 2005;112:848-854.
22. Sturmer J, Gloor B, J TH. Wie sehen glaukomgesichtsfelder wirklich aus? *Klin Mbl Augenheilk* 1984;184:390-393.
23. Sturmer J. What do glaucomatous visual fields really look like in fine-grid computerized profile perimetry? *Dev Ophthalmol* 1985;12:1-47.
24. Westcott MC, McNaught AI, Crabb DP, Fitzke FW, Hitchings RA. High spatial resolution automated perimetry in glaucoma. *Br J Ophthalmol* 1997;81:452-459.
25. Airaksinen, Heijl A. Visual field and retinal fibre layer in early glaucoma after optic disc haemorrhage. *Acta Ophthalmol* 1983;61:186-194.
26. Zhu S, Wu Y, Mumford D. Filters, random fields and maximum entropy (FRAME): towards a unified theory for texture modeling. *Intl J Comp Vision* 1998;27:107-126.
27. Chauhan BC, Tompkins JD, LeBlanc RP, McCormick TA. Characteristics of frequency-of-seeing curves in normal subjects, patients with suspected glaucoma, and patients with glaucoma. *Invest Ophthalmol Vis Sci* 1993;34:3534-3540.
28. Wall M, Kutzko KE, Chauhan BC. Variability in patients with glaucomatous visual field damage is reduced using size V stimuli. *Invest Ophthalmol Vis Sci* 1997;38:426-435.
29. Maddess T, Severt WL, Stange G. Comparison of three tests using the frequency doubling illusion to diagnose glaucoma. *Clin Experiment Ophthalmol* 2001;29:359-367.
30. Artes PH, Chauhan BC. Signal/noise analysis to compare tests for measuring visual field loss and its progression. *Invest Ophthalmol Vis Sci* 2009;50:4700-4708.
31. Heijl A. Perimetric point density and detection of glaucomatous visual field loss. *Acta Ophthalmol* 1993;71:445-450.
32. Weber J, Dobek K. What is the most suitable grid for computer perimetry in glaucoma patients? *Ophthalmologica* 1986;192:88-96.
33. Wall M, Woodward KR, Doyle CK, Zamba G. The effective dynamic ranges of standard automated perimetry sizes III and V and motion and matrix perimetry. *Arch Ophthalmol* 2010;128:570-576.
34. Campbell FW, Green DG. Optical and retinal factors affecting visual resolution. *J Physiol* 1965;181:576-593.
35. Artes PH, Nicolela MT, McCormick TA, LeBlanc RP, Chauhan BC. Effects of blur and repeated testing on sensitivity estimates with frequency doubling perimetry. *Invest Ophthalmol Vis Sci* 2003;44:646-652.
36. Petersen DP, Middleton D. Sampling and reconstruction of wave-number-limited functions in n-dimensional euclidian spaces. *Inform Control* 1962;5:279-323.