A New Rapid Threshold Algorithm for Short-Wavelength Automated Perimetry

Boel Bengtsson

PURPOSE. To develop and test a short and reliable visual field threshold program for the early detection of glaucomatous visual field loss, by adapting the Swedish interactive test algorithm (SITA) to short-wavelength-automated-perimetry (SWAP).

METHODS. Computer simulations were performed to test the accuracy of several versions of SITA SWAP, and to optimize speed versus reliability. The selected SITA SWAP version was evaluated and compared with the older Full Threshold SWAP and Fastpac SWAP programs in 41 patients with glaucoma and normal subjects.

RESULTS. Average test time was 3.6 minutes for SITA SWAP, 11.8 minutes for Full Threshold SWAP, and 7.7 minutes for Fastpac SWAP, differences were significant at \( P < 0.0001 \). Mean threshold reproducibility, calculated as absolute difference between two tests, did not differ significantly between programs and was 2.4 dB for SITA, 2.3 dB for Full Threshold, and 2.4 dB for Fastpac SWAP. Simultaneous comparison showed significant differences in threshold sensitivity, \( P = 0.0234 \): SITA SWAP showed highest sensitivity, 21.6 dB on average, compared with both Full Threshold SWAP and Fastpac SWAP with a mean sensitivity of 17.3 and 17.8 dB, respectively.

CONCLUSIONS. SITA SWAP was much faster than the older SWAP strategies, and reproducibility did not differ. This implies that SITA SWAP could become a clinically useful method for the detection of early glaucoma. SWAP tests may also be applicable in larger groups of patients because of the increased dynamic range. (Invest Ophthalmol Vis Sci. 2005;46:1388–1394) DOI: 10.1167/iovs.02-0169

Short-wavelength perimetry (SWAP) applies a blue stimulus, Goldmann size V, on an intense yellow background. The currently accepted clinical method, as incorporated in the Humphrey Field Analyzer (HFA; Zeiss-Humphrey Systems, Dublin, CA), utilizes a Goldmann size V narrow-band blue stimulus with a peak transmission of 440 nm presented on a 100-cd/m² yellow background.1 SWAP can detect glaucomatous visual field loss before conventional white-on-white (WW) perimetry.2–6 Whether this advantage is based on selective losses of a subpopulation of ganglion cells in glaucoma or on reduced redundancy is debatable—that is, whether SWAP responses are simply derived from a subset of ganglion cells that are selectively damaged by glaucoma, or, alternatively, whether because SWAP tests only one type of ganglion cells, other cell types that may still be functional in a given test location therefore cannot contribute to detection of the blue-on-yellow target.7 There are limitations associated with the SWAP technique, however.8 SWAP results are affected by age-related yellowing of the lens9,10 and by cataract.11,12

SWAP has been reported to require 15% to 17% more time than WW perimetry using the same threshold strategies.13 The longer test time with SWAP compared with conventional WW perimetry is a practical disadvantage.

Variability in threshold estimates is greater with SWAP than with WW.13–16 Intersubject variability in normal subjects, which forms the base for normal limits of threshold deviations, is also larger with SWAP13,15,17 than with conventional WW perimetry. Hence, the normal limits for SWAP thresholds are wider than those for WW thresholds, resulting in less sensitivity for the same decibel depression with SWAP perimetry and hence less sensitive probability maps18 (Fig. 1).

The SITA Standard and Fast programs were originally developed for WW perimetry with the goal of considerably shortening test times without losing accuracy, compared with the older test programs they were designed to replace: the Full Threshold and Fastpac strategies, respectively (Zeiss-Humphrey Systems).19,20 SITA Standard tests are approximately 50% shorter than the older Full Threshold test, and SITA Fast is approximately 50% shorter than Fastpac tests, while showing similar reproductibility.21–25 Threshold sensitivity is higher with the SITA programs (i.e., the dynamic range of stimulus intensity is increased). Intersubject variability is smaller and threshold normal limits are narrower than with the older threshold programs.26,27 In glaucomatous eyes, slightly more test points show significantly depressed threshold deviations with SITA, and sensitivity to glaucomatous visual field damage is at least as good as with the older standard programs.24,28

The first purpose of the present study was to modify the SITA algorithm to SWAP, to obtain a fast and reliable test for early detection of glaucomatous field defects. A second purpose was to conduct a pilot clinical test with the new algorithm.

MATERIALS AND METHODS

SITA SWAP Algorithm

Threshold calculations in SITA are based on a model of the visual field.13 The model includes age-corrected normal thresholds, inter-point correlations, and frequency-of-seeing (FOS) curves. All these elements differ for different sizes and types of stimuli. Thus, the original SITA model developed for WW perimetry had to be modified for SWAP. Age-corrected normal thresholds and age slopes were retrieved from the normal database of SWAP fields used for calculation of the commercially available Humphrey Statpac (Zeiss-Humphrey Systems). The database consisted of 392 eyes of 392 subjects, with ages ranging from 20 to 80 years, collected at 10 different centers. Subjects having intraocular pressure more than 22 mm Hg in either eye; occluded angles; visual acuity less than 20/50; refractive error 5 D sphere

From the Department of Ophthalmology, Malmö University Hospital, Malmö, Sweden.

Supported by Zeiss-Humphrey Systems, Dublin, California.

Submitted for publication February 19, 2002; revised June 20, 2002; accepted July 9, 2002.

Commercial relationships policy: F.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Corresponding author: Boel Bengtsson, Department of Ophthalmology, Malmö University Hospital, SE-20502 Malmö, Sweden; boel.bengtsson@oftal.mas.lu.se.

Copyright © Association for Research in Vision and Ophthalmology
or 2.5 D cylinder or more; discs with known or suspected disease; reproducible field defects explicable by ocular status or history; previous or current eye disease; serious eye trauma or any intraocular surgery; abnormal pupil; a history of amblyopia, diabetes, or lupus; a history of treatment with medications that may be expected to affect the visual field; inability to undergo a visual field test; or a known blue-yellow color defect were not included (Patella M, Zeiss-Humphrey Systems, personal communication, 1995). Interpoint correlations across the field were calculated with the manufacturer-provided normal SWAP database, and clinical SWAP fields obtained in patients with mild to moderate field loss as determined by WW perimetry and with typical glaucomatous disc appearance.

Variability in threshold estimates has an important role in the SITA visual field model, and SWAP threshold variability was determined by measuring FOS curves in normal and glaucomatous eyes. The threshold is defined as the stimulus intensity that has a 50% probability to be perceived. The slope characteristic, $\sigma$, is important. A small $\sigma$ means that the FOS curve is steep, the variability in the threshold estimate is small, and thus that the threshold is well defined. A flat curve has a large $\sigma$ indicating large variability, and thresholds are not as well defined. It is known from WW perimetry that test-retest variability and FOS $\sigma$ are larger at defective locations than at normal ones.29–31 This is true also for SWAP, but the SWAP FOS $\sigma$ is also twice as large as the WW $\sigma$. Thus, we found $\sigma$ of 1.79 dB for SWAP and 0.86 dB for WW perimetry in a study involving a small number of normal subjects ($n = 15$) and patients with glaucoma ($n = 18$).15 Results that we later confirmed in 35 additional subjects (Bengtsson B, unpublished results, 2000). The SWAP FOS curves were incorporated in the SWAP field model.

The SITA test procedure is similar to older visual field testing in the sense that stimulus intensities are altered up and down in staircases, but all answers to stimuli, both seen and not seen, are added to the threshold model, which is continuously recalculated during the actual test. All answers are also used for calculation of variability in threshold estimates. Test time is saved without decreasing the quality of threshold estimates by interrupting stimulus staircases sooner at test points where variability in threshold estimates is small.19

**Computer Simulations**

We modified the simulator$^{19}$ that was used to develop the first SITA for WW perimetry. The simulator could execute many different versions of SITA SWAP, short and long ones, and with different accuracy in threshold estimates. Threshold data obtained by real 24-2 Full Threshold SWAP tests in 80 normal eyes of 80 healthy volunteers with normal discs, visual acuity of 20/50 or better, and refraction less than ±2.5 D and of 25 eyes of 25 patients with visual field defects and discs with glaucomatous appearance were used as the input data. These were regarded as true thresholds. The patients’ response characteristics were determined by empirically derived SWAP FOS curves, plus random frequencies of false answers derived from distributions of 130 normal and abnormal fields.$^{52}$ The accuracy of simulated fields was calculated as the absolute difference between the true threshold, defined by the input field data, and the simulated threshold. Test duration was expressed as the total number of questions per test. Simulations were used to tune the accuracy of thresholds versus test time.

Simulations were performed with different stimulus step sizes. It would have been possible to let the SITA SWAP threshold calculation model determine the intensity of stimuli. This was tried early in the development of SITA WW and resulted in increased test duration without better quality of the threshold estimate.$^{53}$ Therefore, this feature was never used when developing SITA SWAP. Instead, we tested various combinations of stimulus staircases with fixed step sizes. During the simulations, all other test parameters were kept constant to isolate the effect of step sizes. In traditional staircase methods for the estimation of thresholds, large steps resulted in cruder threshold estimates than small steps.$^{54}$ Larger steps, however, could be expected to be more efficient in SWAP than in WW perimetry because of the flatter FOS curves. Results from the SITA SWAP simulations of glaucomatous fields showed that 3-dB steps gave accuracy similar to 4-dB steps (Fig. 2).

The Error-Related Factor (ERF)$^{19}$ is a feature in SITA that is used to interrupt stimulus staircases when a certain confidence in the threshold estimate is reached. The confidence is based on the variability in the threshold estimate and also on the level of threshold sensitivity. In the final version of SITA SWAP, the ERF was fixed, but during development the magnitude of ERF was adjusted to optimize accuracy versus test duration. Small ERFs means that the variability in the threshold estimate is small and the sensitivity level is high. Simulations were performed using 4-dB steps while varying ERFs. Seven SITA versions with different ERFs were tested in simulations. Small ERFs generally produced results with somewhat higher accuracy, but more questions were needed to reach the confidence limit than SITA versions with large ERFs. The differences were rather small, however; means of absolute differences between true and simulated thresholds ranged from 1.91 to 2.07 dB between the different versions of SITA SWAP. The number of questions per test ranged from 174 to 210. Four of these versions, which showed high efficiency defined as absolute difference $\times$ number of questions per test, were tested on patients with glaucoma.

**Real-Time Testing of Experimental SITA SWAP Versions**

One eye of each of 10 patients was tested twice with four different versions of SITA SWAP at four separate visits. Thus, two tests were obtained at each visit. The four SITA SWAP versions, A, B, C, and D, applied 4-dB steps until a first reversal in the staircase. One version, D, in addition applied 2-dB steps after the first reversal at those points located within 12° of eccentricity (Table 1). ERFs were altered. The
The smallest ERF was applied in the C version. With such small ERFs the required confidence limit set to interrupt stimulus sequences was not reached at any test point. Thus, full staircases were performed at all 54 test point locations included in the 24-2 test point pattern. The SITA versions applying higher ERFs showed a larger number of test points with interrupted staircases (Table 1).

Accuracy is unknown in real-time testing. Instead, we relied on reproducibility, calculated as the absolute difference between repeated point-wise thresholds for each subject and each test program. Test time was computed as the average of the two tests obtained with each program. The results suggested that SITA SWAP version D had the best combination of accuracy and test time (Fig. 3, Table 1).

### Pilot Study

The selected version of SITA SWAP (D) was tested and compared with Full Threshold SWAP and Fastpac SWAP. The 24-2 test point pattern of the HFA was used for all SWAP tests. Thirty-one eyes of 31 patients with glaucoma and 10 eyes of 10 healthy volunteers were studied. The mean age of the patients with glaucoma was 72 years, ranging from 59 to 82, whereas the mean age was 41 years in the normal subjects, ranging from 33 to 55 years. Twenty-four women, 7 of them normal, and 17 men, 3 of them normal, were included. In the glaucoma group, all stages of the disease were represented as measured by the mean deviation found with SITA Fast (Fig. 4), but with an emphasis on eyes with mild damage, as the purpose of SITA SWAP is to enable early diagnosis. All subjects performed a total of six SWAP tests, two each with SITA version D, plus the Full threshold or Fastpac programs at three separate visits. One SITA Fast WW test was obtained at one of the visits when the SITA SWAP tests were taken, but it was administered after the SWAP tests. The test order of programs was randomized so that approximately one third was tested first with each SWAP strategy.

All subjects, both healthy volunteers and patients with glaucoma gave informed consent. The Committee for Research Ethics at Lund University approved the test protocol and the tenets of the Declaration of Helsinki were observed.

### Analyses

Average test time, mean sensitivity, and reproducibility, calculated as the absolute difference between thresholds of the two tests with each SWAP threshold program, were computed for all subjects. Differences between the three SWAP programs were compared by ANOVA and Scheffe’s test for simultaneous comparisons.

### RESULTS

Test times were considerably shorter with SITA SWAP than with both Fastpac and Full Threshold SWAP (Fig. 5, top). Mean test time with SITA SWAP was 3.6 minutes; with Fastpac SWAP 7.7 minutes; and with Full Threshold SWAP 11.8 minutes, on average. The differences between all programs were significant ($P < 0.0001$).

Threshold reproducibility did not differ between the three strategies, $P = 0.553$ (Fig. 5, middle). The mean absolute difference was 2.4 dB for SITA SWAP, 2.4 dB for Fastpac SWAP, and 2.3 dB for Full Threshold SWAP.

### Table 1. Accuracy and Test Time of the Four SITA SWAP Versions

<table>
<thead>
<tr>
<th>SITA Version</th>
<th>Stimulus Step (dB)</th>
<th>ERF</th>
<th>Threshold Reproducibility (difference in dB)</th>
<th>Average Test Time (min)</th>
<th>Efficiency (Reproducibility × Test Time)</th>
<th>Number of Test Point Locations with Staircases Interrupted by ERF (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4</td>
<td>1.49</td>
<td>3.24</td>
<td>4.06</td>
<td>13.54</td>
<td>28</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>1.29</td>
<td>3.16</td>
<td>4.24</td>
<td>13.72</td>
<td>18</td>
</tr>
<tr>
<td>C</td>
<td>4</td>
<td>0.89</td>
<td>2.55</td>
<td>5.94</td>
<td>15.11</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>4 and 4-2</td>
<td>1.29</td>
<td>2.75</td>
<td>3.58</td>
<td>9.97</td>
<td>16</td>
</tr>
</tbody>
</table>

Downloaded From: http://iovs.arvojournals.org/pdfaccess.ashx?url=/data/journals/iovs/932918/ on 12/14/2017
Mean sensitivity was higher with SITA SWAP (21.6 dB) than with Fastpac SWAP and Full Threshold SWAP (17.8 and 17.3 dB, respectively; Fig. 5, bottom). The difference between SITA SWAP and Fastpac SWAP was significant \((P = 0.026)\) and also between SITA SWAP and Full Threshold SWAP \((P = 0.012)\). There was no significant difference in sensitivity between Fastpac and Full Threshold \((P = 0.774)\).

**DISCUSSION**

The SITA algorithm applies Bayesian statistics for maximum a priori (MAP) likelihood calculation of thresholds. MAP calculation is an established mathematical statistical method, and was described by Watson and Pelli\(^{35}\) for threshold estimation in experimental psychophysical tests. King-Smith et al.\(^{36}\) later developed their method for perimetry, which was adapted later to SWAP by Turpin et al.\(^{37}\) The SITA algorithm, however, is based on the work in studies by Olsson et al.\(^{38}\), Olsson,\(^{39}\) and Olsson and Rootzén\(^{40}\) and was designed not only to estimate thresholds but also to save test time by stopping testing at locations when the variability in the threshold estimate is small enough according to predetermined rules.\(^{19}\) SITA also includes two other time-saving features: a new way to estimate false-positive and false-negative answers\(^{41,42}\) and improved pacing of the test.\(^{19}\)

The selected version of SITA SWAP showed much shorter test times than both older SWAP programs. This was one of the main goals when modifying the SITA algorithm for SWAP testing. Recent development of threshold strategies has focused on shorter test times.\(^{19,20,43,44}\) The long test duration using older SWAP methods has been one of the major obstacles for using SWAP in clinical settings. In clinical work, SWAP has therefore been used only in a minority of patients with glaucoma or suspected glaucoma.

Reproducibility of threshold estimates did not differ between the three strategies. In real-time testing, the clinician has to rely on reproducibility instead of accuracy but reliable threshold algorithms are expected to produce acceptable reproducibility. It should be remembered, however, that crude tests, such as screening programs with only two outcomes, seen or not seen, show higher reproducibility than threshold tests with a wide range of outcomes. Our comparison, however, included different, but similar threshold programs.

**FIGURE 3.** Four different versions of SITA SWAP were tested on 10 patients with glaucoma. The C version required the longest test time \((top)\) and showed the best reproducibility \((middle)\). The D version had the second best reproducibility and shortest average test time. Efficiency, calculated as difference between thresholds \(\times\) test time, was highest with version D \((bottom)\).

**FIGURE 4.** Distributions of mean deviations (MD) of SITA Fast WW tests obtained in the 31 patients with glaucoma. Most patients had mild defects \((i.e.,\text{ MD close to 0})\).
Accuracy can only be calculated when a comparison with true thresholds is possible, as in simulations. This is one important reason to use simulations when developing new threshold programs. Another advantage is that single test parameters, such as step size or ERF, can be varied and evaluated one at the time in a large number of fields much more rapidly than with real-time measurements.

With SITA SWAP average threshold sensitivity increased by approximately 4 dB compared with the older SWAP programs. This is in agreement with our earlier results comparing WW SITA with WW Full Threshold,\textsuperscript{20–22} and also with the results of the Turpin et al.,\textsuperscript{37} who found approximately a 3-dB difference between the SWAP Full Threshold and the method of King-Smith et al.,\textsuperscript{36} adapted for SWAP. The higher sensitivity is positive, thus increasing dynamic range. To some extent, the higher sensitivity is due to differences in the definition of thresholds. Bayesian MAP algorithms define the threshold as the stimulus intensity that has a 50% probability to be seen, whereas the older strategies define the threshold as the stimulus intensity last seen at the end of the staircases. The absence of visual fatigue\textsuperscript{45,46} also contributes to the higher sensitivity of the shorter SITA tests compared with more time-consuming older tests. When developing the SITA programs for WW perimetry, we found a positive correlation between differences in mean sensitivity and differences in test length,\textsuperscript{20} with shorter tests being associated with higher sensitivity.

The difference in threshold sensitivity between test programs has a large influence on the gray-scale representations of such values. This makes a comparison of sensitivity to glaucomatous field loss very difficult when only raw threshold data and resultant gray-scale representations are available. More accurate evaluations should take age-corrected normal thresholds and normal limits for such values into consideration. Such normal thresholds are not available at the present time. A subjective evaluation of the appearance of threshold gray scales, however, suggests that SITA SWAP would show glaucomatous field loss similar to that found with the older SWAP programs. One example is shown in Figure 6. It could be expected that narrower normal threshold limits and more sensitive probability maps as found with WW SITA compared with WW Full Threshold\textsuperscript{26} also will be found with SITA SWAP, but this remains to be tested. However, an initial analysis was performed to calculate preliminary age-corrected thresholds and empirically derived normal limits at the \( P < 0.1 \) and \( P < 0.05 \) levels for comparison between strategies using the 20 tests obtained with each of the three SWAP strategies in the 10 normal subjects. These crudely calculated limits were considerably wider for Fastpac SWAP and Full Threshold SWAP than for SITA SWAP. It therefore seems likely that SITA SWAP probability maps may become more sensitive to early glaucoma loss than the older SWAP programs.

SITA SWAP was developed to become a fast and reliable clinical method for the detection of early glaucomatous visual field loss in patients without cataract. Considerably reduced test time, increased dynamic range and perhaps also smaller intersubject variability imply that SITA SWAP could become such a test, but further testing is necessary. Definite comparisons cannot be made until normal age-corrected thresholds and normal threshold limits have been determined.

**Figure 5.** All subjects had shorter test times with the tested SITA SWAP version than with Fastpac and Full Threshold SWAP (top), whereas reproducibility did not differ between algorithms (middle). Mean sensitivity was always higher with SITA SWAP (bottom).
Full Threshold
Test time: 12:50

Fastpac
Test time: 7:50

SITA
Test time: 3:56

Stimulus: V, Blue
Background: Yellow

Stimulus: V, Blue
Background: Yellow

Stimulus: V, Blue
Background: Yellow

Figure 6. Three SWAP tests. Full Threshold (left), Fastpac (middle), and SITA (right), in the same eye. Comparison of gray scales is difficult because of differences in the height of the field between SITA and the other strategies. Fastpac and Full Threshold SWAP tests appear darker, but the SITA field displays paracentral and nasal defects. A good comparison should include age-normal values and normal limits for each program, as included in the probability maps.

Acknowledgments

The author thanks Buck Cunningham, Software Engineer, Zeiss-Humphrey Systems, Dublin, CA, for programming the experimental SITA SWAP, and Jonny Olsson, PhD, for modifying the SITA calculation model and the computer simulator.

References


