Comparison of unifocal, flicker, and multifocal pupil perimetry methods in healthy adults

Brendan L. Portengen Ophthalmology Department, University Medical Center Utrecht, The Netherlands
Experimental Psychology, Helmholtz Institute, Utrecht University, The Netherlands

Giorgio L. Porro Ophthalmology Department, University Medical Center Utrecht, The Netherlands

Saskia M. Imhof Ophthalmology Department, University Medical Center Utrecht, The Netherlands

Marnix Naber Experimental Psychology, Helmholtz Institute, Utrecht University, The Netherlands

To this day, the most popular method of choice for testing visual field defects (VFDs) is subjective standard automated perimetry. However, a need has arisen for an objective, and less time-consuming method. Pupil perimetry (PP), which uses pupil responses to onsets of bright stimuli as indications of visual sensitivity, fulfills these requirements. It is currently unclear which PP method most accurately detects VFDs. Hence, the purpose of this study is to compare three PP methods for measuring pupil responsiveness.

Unifocal (UPP), flicker (FPP), and multifocal PP (MPP) were compared by monocularly testing the inner 60 degrees of vision at 44 wedge-shaped locations. The visual field (VF) sensitivity of 18 healthy adult participants (mean age and SD 23.7 ± 3.0 years) was assessed, each under three different artificially simulated scotomas for approximately 4.5 minutes each (i.e. stimulus was not or only partially present) conditions: quadrantanopia, a 20-, and 10-degree diameter scotoma.

Stimuli that were fully present on the screen evoked strongest, partially present stimuli evoked weaker, and absent stimuli evoked the weakest pupil responses in all methods. However, the pupil responses in FPP showed stronger discriminative power for present versus absent trials (median *d*-prime = 6.26 ± 2.49, area under the curve [AUC] = 1.0 ± 0) and MPP performed better for fully present versus partially present trials (median *d*-prime = 1.19 ± 0.62, AUC = 0.80 ± 0.11).

We conducted the first in-depth comparison of three PP methods. Gaze-contingent FPP had best discriminative power for large (absolute) scotomas, whereas MPP performed slightly better with small (relative) scotomas.

Introduction

To this day, the method of choice for clinically testing the visual field (VF) is standard automated perimetry (SAP). Current SAP devices (e.g. Humphrey Field Analyzer [HFA] and Octopus perimeter) systematically measure VF loss by (1) restricting head movement with a forehead-chinrest, (2) asking patients to fixate on a central target, and (3) give a motor response when a visual change (usually a temporary increase in luminance or change in color) is shown at one of a number of (typically approximately 54–76) locations across the VF. This procedure of SAP brings along several shortcomings; testing is subjective as it relies on introspective reports due to its psychophysical nature, strict fixation is required, observers need to exert prolonged attention, and learning effects and incorrect motor responses distort measurements. In addition, poor reproducibility has been described for SAP (test-retest variability of single threshold estimates approximating the dynamic range of the instrument; Artes, Iwase, Ohno, Kitazawa, & Chauhan, 2002; Piltz & Starita, 1990). The test-retest variability likely stems from the small stimuli used in SAP and their displacement due to fixational jitter or microsaccades, and learning and fatigue effects (Maddess, 2014;

Citation: Portengen, B. L., Porro, G. L., Imhof, S. M., & Naber, M. (2022). Comparison of unifocal, flicker, and multifocal pupil perimetry methods in healthy adults. *Journal of Vision*, *22*(9):7, 1–13, https://doi.org/10.1167/jov.22.9.7.

https://doi.org/10.1167/jov.22.9.7 Received March 29, 2022; published August 23, 2022 ISSN 1534-7362

Numata, Maddess, Matsumoto, Okuyama, Hashimoto, Nomoto, & Shimomura, 2017; Wall, Woodward, Doyle, & Artes, 2009).

Alternative perimetry methods utilize measurements like visual evoked potentials (i.e. measurement of electrophysiological responses with electrodes positioned on the scalp near the occipital bone), saccadic vector optokinetic perimetry (i.e. saccade measurement in response to visual cues with eye tracking technology), preferential looking responses, saccadic response times, behavioral VF tests, and pupillometry (Allen, Slater, Proffitt, Quarton, & Pelah, 2012; Gestefeld, Grillini, Marsman, & Cornelissen, 2020; Harding, Spencer, Wld, Conway, & Bohn, 2002; Koenraads, Braun, van der Linden, Imhof, & Porro, 2015; Pel, van Beijsterveld, Thepass, & van der Steen, 2013). Here, we specifically focus on pupil perimetry (PP), which emerges as a relatively young and unpursued, but also promising method. Although scientists claim to have improved this form of perimetry, few methodological approaches currently exist and have not yet been compared. Therefore, it is currently unknown how the pupil can best be used as a measure of visual sensitivity.

Optimization strategies so far limited their explorative scope to changing the number of stimuli shown simultaneously across the VF (spatial sparseness) and the frequency of presentations within a certain time window (sparseness of events). In general, three distinct methods that vary across these two factors can be discerned: (1) unifocal PP (UPP), which consists of a single stimulus presentation (i.e. high spatial sparseness) shown for a relatively long period of time (i.e. high sparseness of events; e.g. Kardon, Kirkali, & Thompson, 1991); (2) flicker PP (FPP), recently developed by our laboratory, consisting of a single flickering stimulus presentation (i.e. high spatial sparseness) at gaze-contingent locations, which allows for repeated and precise retinotopic stimulation (i.e. low sparseness of events with regard to the number of luminance changes), circumventing noise and fixation problems that typically occur in slow presentation paradigms (Naber, Roelofzen, Fracasso, Bergsma, van Genderen, Porro, Dumoulin, & van der Schouw, 2018; Portengen, Roelofzen, Porro, Imhof, Fracasso, & Naber, 2021); and (3) multifocal PP (MPP), showing multiple stimuli in parallel for relatively long durations (i.e. low spatial sparseness and high sparseness of events; e.g. Maddess, Bedford, Goh, & James, 2009; Tan, Kondo, Sato, Kondo, & Miyaka, 2001; Wilhelm, Neitzel, Wilhelm, Beuel, Lüdtke, Kretschmann, & Zrenner, 2000); see Figure 1. FPP shows very promising results for application in neurologically impaired subjects affected by cerebral visual impairment on account of its gaze-contingent stimulus presentation with multiple measurements in a short time period (Naber et al., 2018), but a comparison between different PP methods has not yet been performed.

The aim of this study consists of the comparison of sensitivities and specificities across three PP methods.

Methods

Participants

All participants (12 women and 6 men, mean age and SD 23.7 ± 3.0 years) comprised of students and staff of the Psychology Department of Utrecht University with Dutch nationality and Caucasian ethnicity (as observed by the experimenters). All reported having normal uncorrected or corrected visual acuity and no visual or neurological disorders. The experiment conformed to the ethical principles of the Declaration of Helsinki and was preregistered and approved by the local ethical committee of the University Utrecht (approval number: FETC19-006). Participants received (financial) reimbursement for participation and travel and gave informed written consent before the experiment. They were unaware of the purpose of the experiment and were only told that the eye-tracker measured their eye movements. Participants were debriefed about the purpose of the experiment afterward.

Apparatus and stimuli

Stimuli were generated on a Dell desktop computer with Windows 7 operating system (Microsoft, Redmond, WA, USA), using MATLAB (MathWorks, Natick, MA, USA) with the Psychtoolbox 3 and Eyelink toolbox extensions (Brainard, 1997; Cornelissen, Peters, & Palmer, 2002; Kleiner, Brainard, Pelli, Broussard, Wolf, & Niehorster, 2007; Pelli, 1997). A linearized (gamma-correction factor 2.2) OLED65B8PLA LG (LG Electronics, Seoul, South Korea) presentation monitor displayed the stimuli at a resolution of 1920 by 1080 pixels with a refresh rate of 60 Hz. The screen measured 143 cm in width and 63 cm in height. The participant's viewing distance was held stable at 65 cm with a chin- and forehead rest. Pupil size and gaze was recorded monocularly with an Eyelink 1000 eye-tracker (SR Research, Ontario, Canada; 0.5-degree accuracy of gaze location) placed 40 cm in front of the observer below the monitor. Eye-tracker calibration sessions consisted of the presentation of a five-point calibration grid and lasted approximately 1 minute. Note that the Eyelink tracker software outputs pupil size in arbitrary units rather than absolute pupil diameter in millimeters. The experiment was conducted in a darkened room without ambient light.

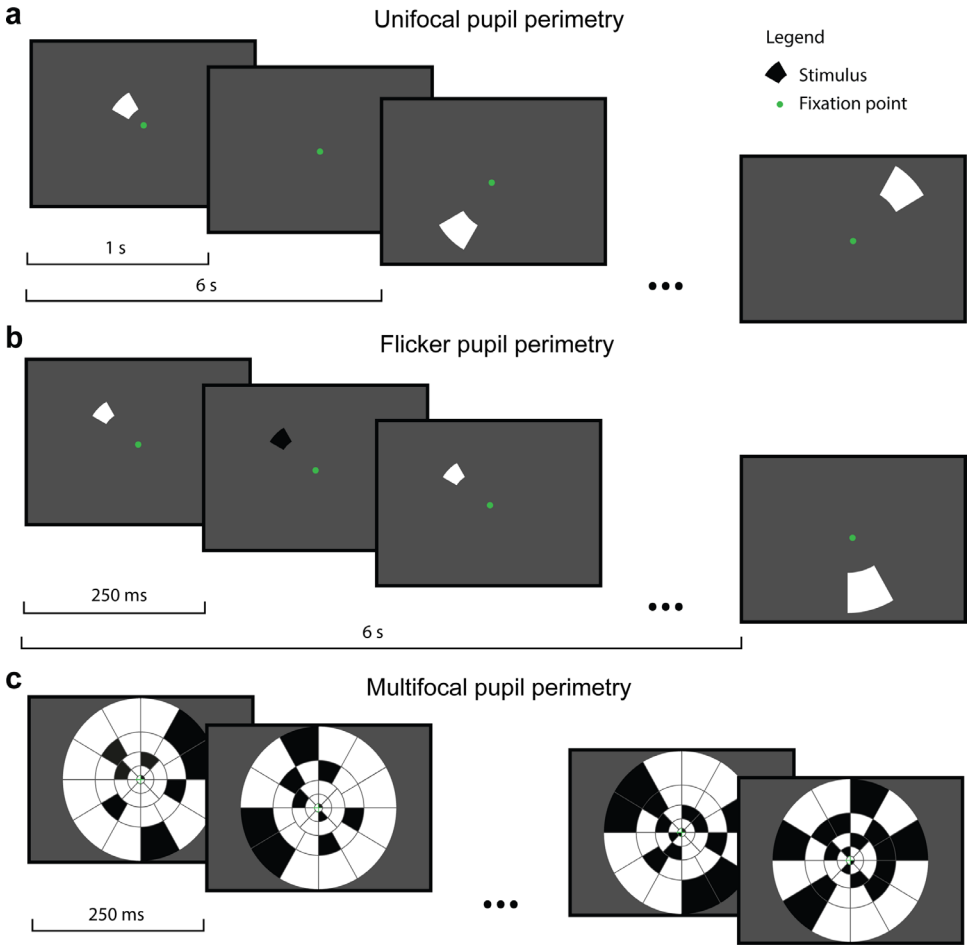

Figure 1. The three pupil perimetry (PP) methods. (**a**) Unifocal PP consisted of a single 1 second duration stimulus presentation followed by a 5 second interval after which another stimulus location was presented. See panel **c** for all stimulus locations used in all methods. (**b**) Flicker PP consisted of black-and-white 2 Hz flickering stimulus presentations for 6 seconds per stimulus location. (**c**) Multifocal PP consisted of the stimulation of several stimulus locations in parallel at any given time point. The temporal stimulation pattern followed an M-sequence calculated with the algorithm of Buračas & Boynton (2002) with 2 Hz refresh rate of stimulus patterns. This ensured best statistical independence across wedges, resulting in the most precise weighting of contributions per wedge to the pupil responses during the analysis. The stimuli for all three methods consisted of 44 wedges of the same size per stimulus location across methods. See Supplementary Videos S1 to S3 for an example of the three methods.

Stimulus paradigms

The three tested methods included UPP, FPP, and MPP, which were tested consecutively in all participants (with short breaks in between the methods and stimulus paradigms; see below) in random order using random permutation. UPP consisted of single, one-by-one stimulus presentations, in the form of white wedges, presented at random order across all 44 stimulus locations, each for a duration of 1 second followed by a 5 second blank screen interval (Figure 1a, Supplementary Video S1-UPP). FPP (Naber et al., 2018) consisted of a 2 Hz flickering wedge with a change in luminance between black and white for a duration of 6 seconds sequentially presented at each individual location (see Figure 1b, Supplementary Video S2-FPP). For MPP, approximately half of the stimulus locations were stimulated in parallel, and the stimulation pattern changed at a rate of 2 Hz for 256 seconds. The interval between stimuli was thus 0.25 seconds, resulting in a total of 1056 stimulus change events per location (see Figure 1c, Supplementary Video S3-MPP). Note that we chose to change temporal and spatial factors of current MPP methods to enable better comparisons with the FPP method. To ensure that the temporal pattern of luminance changes for each wedge location correlated the least as possible to the patterns of the other wedge locations (anticorrelation across wedges improves sensitivity), an m-sequence algorithm was used (Buračas & Boynton, 2002).

To compare the three methods, we assessed scotoma detection accuracy. As such, the observers saw stimuli presented within, at the border, and outside of the areas in which the wedge stimuli were not visible (i.e. not

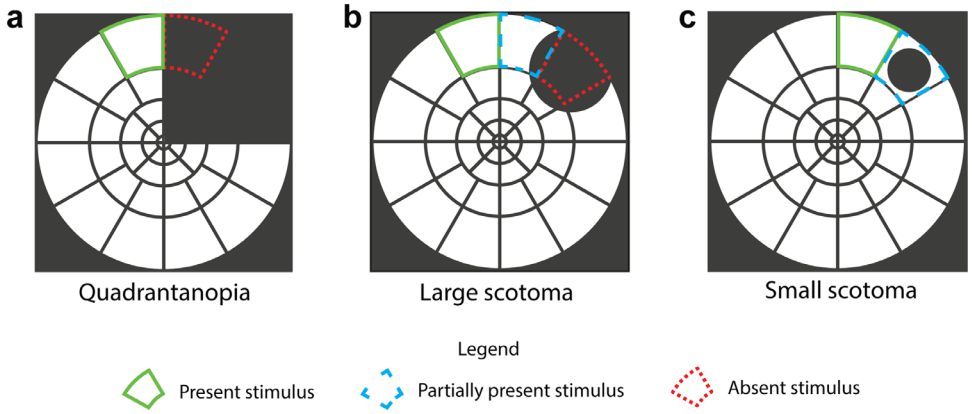

Figure 2. Stimulus location maps per artificial scotoma condition, each consisting of 44 wedges. To compare how well each pupil perimetry method could detect different scotoma types, three artificially simulated visual field defects (aVFDs) were created: quadrantanopia (**a**), relatively large scotoma (**b**), and a relatively small scotoma (**c**). Scotomas were placed either at the upper right or left corner of the visual field. These aVFDs resulted in three distinct wedge visibility conditions: fully present (solid green line), partially present (dashed blue), and absent (dotted red).

shown). We created three scotoma versions with these artificial visual field defects (aVFDs): a stimulus wedge could either be fully present, partially present, or absent (Figure 2). The aVFDs were randomly placed in the upper right or left quadrant of the stimulus map per participant. Simulating VFDs in healthy participants is a known strategy (e.g. Gestefeld et al., 2020) with the following advantages: (i) it allows the exact controlling of which part of the VF is masked and (ii) bypasses the need to burden patients with having to participate in a study that compares stimulus protocols rather than a newly developed diagnostic method.

Stimulus map

Stimulus locations consisted of 44 wedges distributed across five eccentricity rings in the central 60-degree field of vision, both temporally and nasally (see Figures 1c, 2). The stimulus layout closely resembled other multifocal pupil perimetry protocols (Sabeti, James, Carle, Essex, Bell, & Maddess, 2017; Wilhelm et al., 2000). The wedges differed in size per eccentricity ring (radial width = eccentricity$^{1.12}$; in degrees) to adjust for the cortical magnification factor (i.e. stimuli in the fovea are processed by more cortical tissue) and the distribution of photosensitive retinal cells. The stimulus wedges were white (212 cd/m^2) and the background was dark gray (25 cd/m^2 for UPP and FPP, 13 cd/m^2 for MPP to counteract the slightly stronger local contrasts due to the presence of multiple stimuli) rather than black to reduce straylight effects (i.e. to elevate response thresholds of non-stimulated locations; Portengen et al., 2021). A green bull's eye (0.1-degree radius; not shown in the figures) in the center served as a fixation point.

Procedure

Participants were instructed to fixate on the center of the screen but covertly pay attention to the wedges. To ensure that participants paid attention to the stimuli, participants had to press the spacebar in response to the appearance of cues (Naber, Alvarez, & Nakayama, 2013). Cues consisted of a wedge with thin red edges that appeared in approximately 40% of the trials for 0.25 seconds. Participants were tested on varying times of the day. Only the right eye was tested and recorded, leaving the left eye patched. Test duration for each method was 792 seconds (6 second stimulus presentation, 44 stimulus regions, and 3 aVFDs). Total duration of the experiment, including instructions, eye-tracker calibration, and breaks, was between 40 and 60 minutes per participant. All stimuli were presented in a gaze-contingent manner, meaning that the eye tracking software follows the subject's direction of fixation and updates the position of the stimuli on the monitor real-time to reflect changes in the direction of the gaze.

Analysis

Pupil size data were restructured from the continuous recording with an event-related approach using a series of steps. First, blink periods were deleted from continuous data. Blink on- and offsets were detected by setting a speed threshold of three standard deviations (SDs) above the mean. The removed blink periods were interpolated with a cubic method using the interp1 MATLAB function. In the case of UPP, we used the trial start events to window pupil responses to each

trial (and thus to each wedge; every 6 seconds) in 3 second epochs. For FPP, we chose a 5 second epoch between 1 and 6 seconds after stimulus onset, therewith ignoring the initial constriction in the first second that tends to have a divergent and variable amplitude which complicates accurate FFT power estimations. In the case of MPP, we applied an event-related approach, creating 3000 ms epochs per luminance change (every 250 ms). The pupil data were then filtered per trial. Pupil traces from trial start to trial end were saved in a matrix with each row representing a trial and each column representing a timepoint. Pupil sizes were then baseline corrected by subtracting pupil size at stimulus onset. Except for MPP, pupil size was filtered for low frequency noise by subtracting a low-pass Butterworth fit (third order, 0.2 Hz cutoff). This correction allowed comparisons across participants and for FPP it ensured that the 2 Hz signal fluctuated around zero for proper frequency analyses. In addition, pupil size data were filtered to remove high-frequency noise by applying a low-pass Butterworth filter (fifth order, 30 Hz cutoff). UPP and FPP trials were removed if the pupil size variance within a trial crossed a threshold of four SD above the mean. The latter removal procedure was iterated in three loops. Note that for UPP and MPP, the pupil size moves back to baseline before the end of the 3 second epoch duration.

Subsequently, pupil size as a function of time from trial or epoch onset was first normalized across eccentricities. The average pupil traces for trials with stimulations of the largest eccentricity (fifth outer ring) without scotomas served as a baseline and any deviations from its average pupil trace for the other eccentricities were corrected. The pupil sensitivity was determined from the filtered pupil traces in a different manner per perimetry method: for UPP, the pupil constriction amplitude was used as a measure of pupil sensitivity. It was extracted per trial by subtracting minimum pupil size within a 200 to 1200 ms time window after trial onset (i.e. the period a pupil constriction has ended) from the maximum pupil size within a 0 to 500 ms time window after trial onset (i.e. the period a pupil constriction starts). For FPP, pupil oscillation power from a periodogram at 2 Hz served as a measure of pupil sensitivity. Full trial periods of pupil size were each converted to the frequency domain using a Lomb-Scargle algorithm. The convergence and calculation of pupil oscillation power was independent of and thus not affected by individual variability in phase (Naber et al., 2018; Portengen et al., 2021). For MPP, pupil sensitivity was operationalized as the absolute area under the event-related pupil response (ERPR) averaged across all luminance changes per wedge within a time window of 250 to 1500 ms (i.e. the period an ERPR was present and not yet moved back to baseline). For consistency, we reference to all three different manners of pupil measurement calculation as "pupil responsiveness" from now on.

To determine whether pupil responsiveness differed across scotoma types, we performed a repeated measures ANOVA. Two-dimensional pupil sensitivity maps were created with a harmonic spline interpolation to fill the gaps between the centers of the 44 stimulus wedge locations. The performance of each perimetry method was based on how well the method distinguished between present and absent stimuli across trials. The comparison across methods was made with the index *d*-prime (i.e. an index of the discriminability of a signal, given by the separation between the peaks of the probability distributions, defined in z-scores), the area under the curve (AUC) of the receiver operating characteristics (ROC), and the adjusted effect size for small sample sizes (Hedge's *g*). The *d*-prime, and AUC values per participant were compared across the three methods with paired double-sided *t*-tests. Stimulus protocol scripts, data, and analysis files are available on https://osf.io/bqwk8.

Results

We were interested in how the three different pupil perimetry methods differed with respect to how well they detected an aVFD. First, we inspected whether the pupil responded according to our expectations: a relatively fast constriction after stimulus onset and a slower return to baseline for UPP, pupil size oscillations at a rate of approximately 2 Hz for FPP, and significant response to luminance changes per wedge for MPP. These expectations were confirmed (Figure 3a; see Supplementary Figure S1 for responses per scotoma condition).

Next, we checked whether trials with absent and partially absent stimuli evoked weaker pupil responses than trials with present stimuli. Indeed, the pupil responsiveness differed significantly across visibility conditions for all three methods (Figure 3b; see Supplementary Tables S1 and S2, for ANOVA results and post hoc comparisons).

To inspect differences in pupil responsiveness across stimulus locations, we plotted VF sensitivity heatmaps (Figure 4). These maps showed that pupil responses significantly decreased in the scotoma regions, especially for the scotoma types quadrantanopia and large scotoma. The maps of the small scotoma condition contained somewhat increased variability in response (i.e. noise) across the VF. In addition, both the pupil responsiveness in scotoma regions and the amount of noise across all regions appeared to be lowest for the FPP method.

To inspect how well the PP methods dissociated between unstimulated (artificial scotomas) and

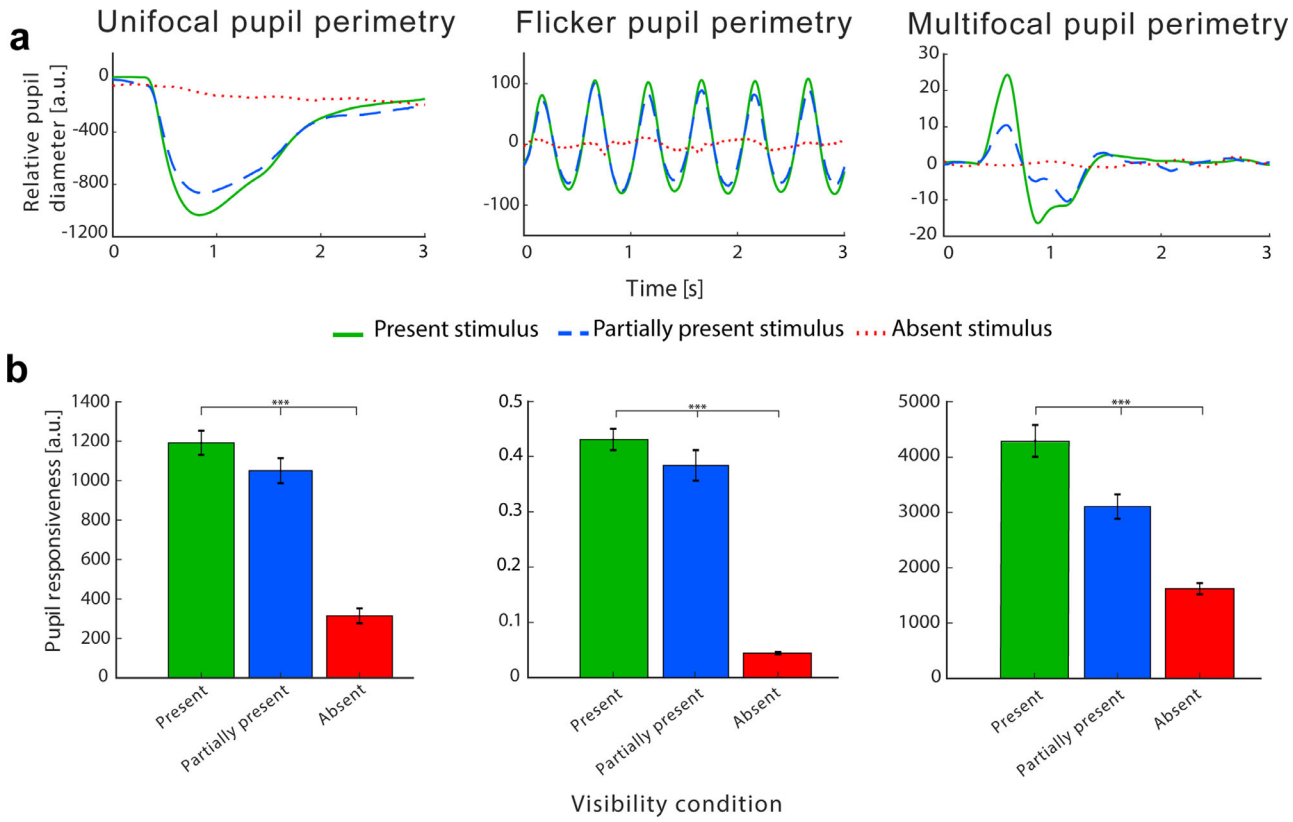

Figure 3. Pattern of pupil responsiveness as a function of time averaged across locations per participant for unifocal pupil perimetry (UPP), flicker pupil perimetry (FPP), and multifocal pupil perimetry (MPP). (**a**) Pupil traces per visibility condition; present (solid green), partially present (dashed blue), and absent (dotted red) stimuli. Note that only the first three seconds of the FPP stimulus duration were depicted to improve comparability across methods. (**b**) Pupil responsiveness per visibility condition (same colors as Figure 3a) per method (panels) averaged across participants with standard errors from the mean. See Figure S1 for results per scotoma condition. Note that the scale of the y-axis differs across methods and between panels **a** and **b** because of the distinct ways the pupil responsiveness is calculated per method (see the Methods section; this does not harm the within-subject comparisons).

stimulated (intact) VF, we created histograms of trial probability as a function of pupil responsiveness for present, partially present, and absent stimulus conditions per PP method (Figure 5a). To quantify the dissociation performance of the PP methods, we calculated *d*-prime, AUC, and Hedge's *g* values per participants, which showed highest sensitivity for the FPP method for present versus absent trials (Figure 5b; median *d*-prime values: UPP = 4.65 ± 1.54, FPP = 6.26 ± 2.49, and MPP = 3.07 ± 1.17; see Table 1) and partially present versus absent trials (see Figure 5b; median *d*-prime values: UPP = 3.73 ± 2.29, FPP = 13.84 ± 6.46, and MPP = 3.14 ± 1.74). Differences were smaller for present versus partially present trials (see Figure 5b; median *d*-prime values: UPP = 0.87 ± 0.94, FPP = 0.78 ± 1.07, and MPP = 1.19 ± 0.62). Statistically comparing the *d*-prime and AUC values per participant across methods (see Figure 5b; see Supplementary Figure S2 for violin plots, and Supplementary Tables S3 and Table S4, for statistics) revealed that FPP produced the least overlapping and most separated pupil sensitivities between absent and present, and partially present versus absent stimulus conditions. MPP performed best for distinguishing present versus partially present trials.

Discussion

This is the first study comparing three different PP methods. From our results, we can conclude that all three PP methods show high discriminative power for differentiating between present and absent stimuli, and between partially present and absent stimuli in healthy adults.

Especially FPP turned out to be qualified to distinguish between present and absent (and partially present) stimuli. One explanation for FPPs high diagnostic accuracy might be that the combination of the single stimulus presentation with an increased number of pupillary measurements in a short time period resulted in multiple, reliable phasic pupil responses (i.e. decreasing the chance of incidental

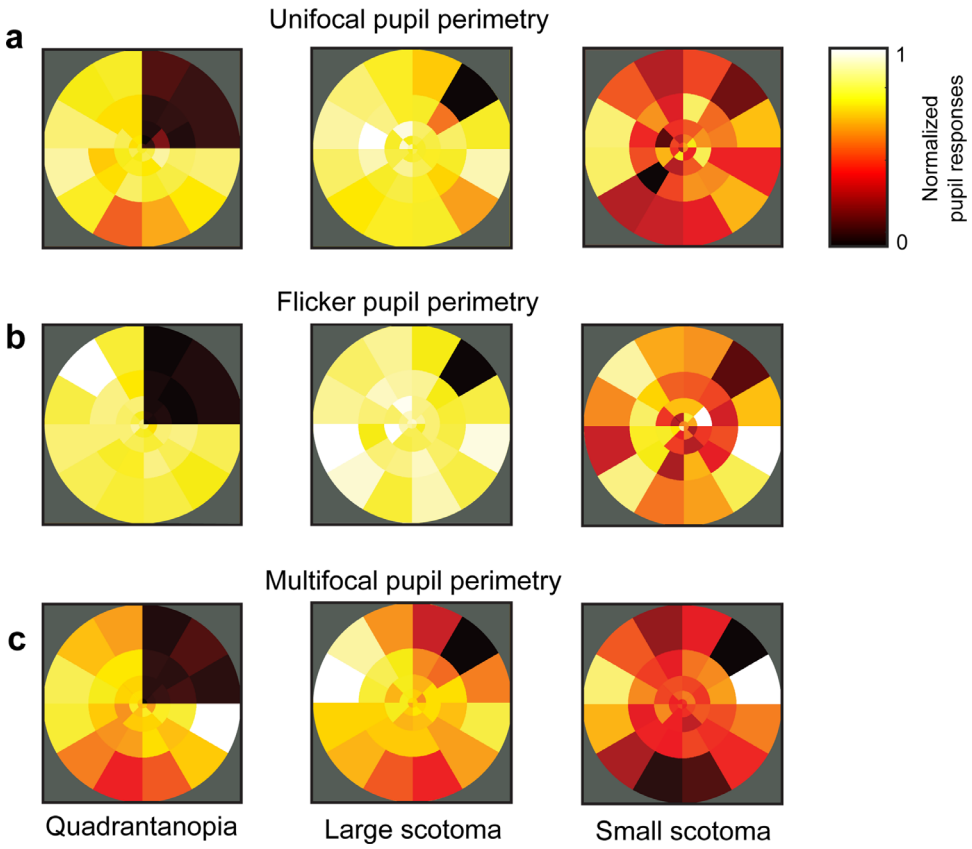

Figure 4. Normalized high-resolution pupil responsiveness visual field heatmaps averaged across all participants per pupil perimetry method (i.e. unifocal [**a**], flicker [**b**], and multifocal [**c**] pupil perimetry) and artificial scotoma condition (left: quadrantanopia, center: large scotoma, and right: small scotoma). Maps of pupil responses from conditions in which artificial scotomas were presented on the left hemifield were horizontally flipped to create this figure. For the quadrantanopia and large scotoma conditions, pupil responses significantly decreased across all methods. The small scotoma condition produced more variance in pupil responsiveness across locations in all methods. Flicker pupil perimetry produced few spurious results in scotoma locations and the least variance across all locations.

pupil fluctuations). These responses could in turn be particularly well suited to distinguish between within-field anisotropies as opposed to looking at average sensitivity across the VF and between damaged and intact VFs in a clinical setting (Naber et al., 2018). Others used flickering stimuli at higher frequencies (i.e. 15 and 30 Hz; James, Kolic, Bedford, & Maddess, 2012; Sabeti, Maddess, Essex, Saikal, James, & Carle, 2014). However, frequencies above 3 to 4 Hz do not evoke the oscillating pupil responses inherent to the flickering method of this study (Naber et al., 2013). The results suggest that a stimulus paradigm with high spatial sparseness and low sparseness of events leads to overall best power in dissociating present, partially present and absent stimuli. The high pupil sensitivity to detect hemianopic and quadrantanopic scotomas due to cortical damage, and glaucoma-caused scotomas, displayed in the first FPP study of Naber et al. (2018), endorse the results found in this study.

Our results showed small between-subject differences for sensitivity measures across visibility conditions and PP methods. Conversely, large individual variation was seen in present versus partially present trials; distinguishing between these conditions remains a challenge when using PP methods (MPP performed only slightly better). This imprecision can partly be explained by the use of large stimulus sizes, which sacrifices spatial precision in the peripheral VF. The presentation of large stimuli is a prerequisite for evoking more reliable pupil responses, but results in coarse sensitivity maps.

It is also not yet possible to dissociate exact VFD locations within a stimulus wedge. To resolve this, a similar stimulus map used by Maddess, Essex, Kolic, Carle, & James (2013), which uses overlapping stimuli shown at different time intervals, or smaller stimuli at more locations like Naber et al. (2018) could be used (with weaker pupil responses as a result). Thus, PP methods are currently more suited for screening purposes than for regular follow-up and monitoring small changes in the VF across time. Conversely, because of the flexible setup of pupil perimetry,

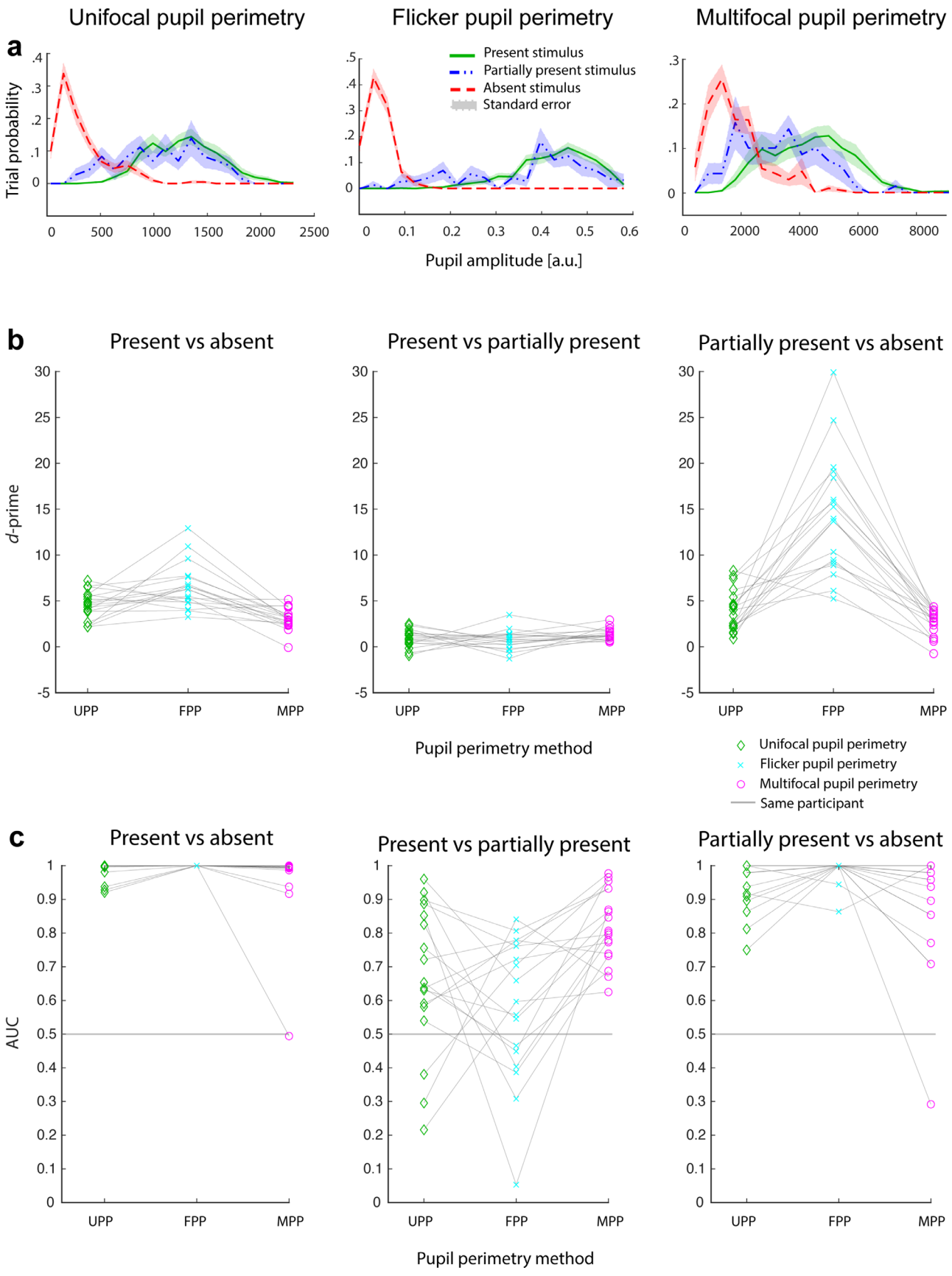

Figure 5. Sensitivity comparison of three pupil perimetry (PP) methods; unifocal (UPP), flicker (FPP), and multifocal pupil perimetry (MPP). (**a**) Shows the number of trials per pupil sensitivity for present (blue continuous line), partially present (green dash and dotted line), and absent (red dashed line) stimuli per PP method (panels). (**b**) Depicts *d*-prime plots per PP method (marker colors) per

→

←
participant (transparent gray lines connecting markers) for present versus absent (left panel), present versus partially present (middle panel), and partially present versus absent (right panel) stimuli; higher *d*-prime values correspond to greater distinctive properties. (**c**) Shows AUC plots per PP method (marker colors) per participant (dotted black lines) for present versus absent (left panel), present versus partially present (middle panel), and partially present versus absent (right panel) stimuli.

			Partially present	Absent
Unifocal PP	**Present**	*AUC* ± *SD*	0.65 ± 0.21	0.99 ± 0.03
		d-prime ± *SD*	0.87 ± 0.94	4.65 ± 1.54
		Hedge's g (CI)	0.60 (0.33; 0.88)	3.01 (2.82; 3.22)
	Partially present	*AUC* ± *SD*		0.98 ± 0.07
		d-prime		3.73 ± 2.29
		Hedge's g (CI)		2.57 (2.15; 3.08)
Flicker PP	**Present**	*AUC* ± *SD*	0.63 ± 0.21	1.0 ± 0.00
		d-prime ± *SD*	0.78 ± 1.07	6.26 ± 2.49
		Hedge's g (CI)	0.57 (0.22; 0.93)	5.05 (4.77; 5.38)
	Partially present	*AUC* ± *SD*		1.0 ± 0.03
		d-prime ± *SD*		13.84 ± 6.46
		Hedge's g (CI)		4.54 (3.78; 5.65)
Multifocal PP	**Present**	*AUC* ± *SD*	0.80 ± 0.11	0.99 ± 0.12
		d-prime ± *SD*	1.19 ± 0.62	3.07 ± 1.17
		Hedge's g (CI)	0.79 (0.54; 1.02)	1.94 (1.79; 2.10)
	Partially present	*AUC* ± *SD*		0.96 ± 0.18
		d-prime ± *SD*		3.14 ± 1.74
		Hedge's g (CI)		1.35 (1.02; 1.72)

Table 1. Comparison of visibility conditions per pupil perimetry method (area under the receiver operating characteristics curve, *d*-prime, and corrected effect size measures through Hedge's *g*). Median values are reported for AUC and *d*-prime. SD = standard deviation, CI = confidence interval.

protocols can easily be interchanged and adjusted. Varying PP protocols could be incorporated for different goals; larger and less stimuli to quickly screen for clinically significant VFDs, and smaller stimuli at more locations to accurately detect small changes during follow-up. Further development could entail automation of a direct diagnostic report and a scotoma edge detection algorithm.

Note, however, that improvements can still be made to the current PP paradigms. Most developments have been reported for MPP (Carle, James, Colic, Essex, & Maddess, 2015; Carle, James, Sabeti, Kolic, Essex, Shean, Jeans, Saikal, Licinio, & Maddess, 2022; Sabeti, James, & Maddess, 2011; Sabeti, James, Essex, & Maddess, 2013; Tan et al., 2001; Wilhelm et al., 2000). Our MPP variant was performed with a relatively high framerate (a possible change in luminance every 250 ms) and long stimulus-on durations and thus differed from state-of-the-art MPP methods in some respects. For example, the method of Wilhelm et al. (2000) involved a scaled honeycomb array and covered 50 degrees of VF, their stimuli were presented with a 50% probability in each test-region, similar to the original electroretinogram (ERG) multifocal method proposed by Sutter (1991; Sutter & Tran, 1992); Tan et al. (2001) created temporally more sparse stimuli by inserting blank frames between frames containing stimuli; Sabeti et al. (2011), and Ho, Wong, Carle, James, Kolic, Maddess, & Goh. (2010) used colored stimuli and a higher presentation frequency, resulting in high temporal sparseness due to short stimulus durations and long inter-stimulus intervals. The most recent MPP method of Carle et al. (2022) features a clustered volley technique, which brings the stimuli closer to each other, and longer interstimulus times than previous iterations, actually making it resemble FPP more with respect to spatiotemporal properties. However, Carle et al.'s MPP method also implements color, luminance balancing (i.e. variance in luminance across stimulus locations), and no black stim-off region. Nonetheless, these improvements can also be implemented in FPP (and UPP), meaning that the here reported differences across PP methods remain valid despite the use of rather basic stimulus paradigms.

It is possible that pupil responses become more sensitive when evoked with fewer stimulus changes per second (e.g. 1 Hz instead of 2 Hz) and a spatial sparseness somewhere in between the range of 1

and approximately half of the 44 locations, as pupil responses seem to be stronger when more stimuli are shown, even at a constant luminance (Castaldi, Pomè, Cicchini, Burr, & Binda, 2021). Although out of the scope of the current study, an optimal spatial and temporal sparseness remains to be found. Nonetheless, the main advantage of endorsing a lower temporal sparseness lies within more data points per trial and consequently shorter testing times.

Although unifocal and flicker PP methods benefit from an attentional cueing paradigm (Binda & Murray, 2015; Einhäuser, 2017; Mathôt & Van der Stigchel, 2015; Naber et al., 2013; Portengen et al., 2021), evidence has been provided that a centrally directed attentional task and covertly directed attention reduces signal quality on multifocal methods (Rosli, Carle, Ho, James, Kolic, Rohan, & Maddess, 2018). This likely stems from a divided attention across multiple simultaneously shown stimuli.

The current study used a dark gray background to suppress the influence of stray light (seen with black backgrounds) and to increase pupil responsiveness (as compared to a lighter gray background; Portengen et al., 2021). This testing method may be improved even more by implementing chromatic properties, such as hue, brightness, and saturation, to strengthen pupil response amplitudes driven by contrasts between those properties and to isolate the retinal opsin, rhodopsin, or melanopsin pathways (Carle et al., 2015; Chibel, Sher, Ben Ner, Mhajna, Achiron, Hajyahia, Skaat, Berchenko, Oberman, Kalter-Leibovici, Freedman, & Rotenstreich, 2016; Maeda, Kelbsch, Straßer, Skorkovská, Peters, Wilhelm, & Wilhelm, 2017). The use of narrow band yellow (around 580 nm) rather than full visible spectrum white (the latter includes blue light) stimuli may reduce blue color-sensitive melanopsin retinal ganglion cell activity and its effect on pupil responses and therewith could contribute to a more accurate diagnosis of VFDs specifically caused by cortical damage (Rosli et al., 2018).

A limitation of the current study is that no normative data from a "no scotoma condition" was used in the analysis, and left versus right VFs per participant may have contained small biases due to temporal versus nasal anisotropies. Although these biases did not hamper the comparison between methods, overall discriminative power could be weaker than when normative data were used. Another limitation pertains to the use of hard edges for the artificially simulated VFDs, which do not accurately represent real world situations with actual VFDs. Although simulating VFDs in healthy participants is an established strategy (e.g. Gestefeld et al., 2020), it does not mimic VFDs entirely. Real scotomas tend to have smooth edges with a gradual gradient from visible to invisible. Due to limitations of the used computer, computing soft edged wedges leads to technical problems, such as slower frame rates. The wedges were created in real-time to ensure a different order of appearance for each participant. In future studies, this could be resolved by creating multiple videos with random presentation orders in advance rather than on-line stimulus buffering. Regardless, several studies showed promising results with PP in more realistic situations, such as detecting the blind spot (Portengen et al., 2021) and testing patients suffering from VFDs (e.g. Carle et al., 2015; Chibel et al., 2016; Kardon, 1992; Maeda et al., 2017; Naber et al., 2018; Rajan, Bremner, & Riordan-Eva, 2002; Schmid, Luedtke, Wilhelm, & Wilhelm, 2005; Skorkovská, Lüdtke, Wilhelm, & Wilhelm, 2009; Skorkovská, Wilhelm, Lüdtke, & Wilhelm, 2009; Tan et al., 2001; Yoshitomi, Matsui, Tanakadate, & Ishikawa, 1999). Future studies testing subjects with VF defects due to neurological impairment will further clarify the role of PP in testing VFs.

As a last point, it is important to stress PPs advantages over SAP. In addition to its high accuracy in detecting artificial scotomas, PP is an objective method for testing VF in a short amount of time (approximately 4 minutes per eye and method). This is comparable to subjective fast SAP methods, such as the Swedish Interactive Testing Algorithm (SITA) 24-2 FAST and Tendency-Oriented Perimetry (TOP). Combined with the minimal cooperation required, this method might have merit for application in young children or neurologically impaired subjects affected by cerebral visual impairment who generally show difficulties in completing an SAP test reliably (Patel, Cumberland, Walters, Russell-Eggitt, Rahi, & OPTIC Study Group, 2015; Wong & Sharpe, 2000). Current alternatives for young or neurologically impaired children are behavioral perimetry tests, such as the behavioral VF (BEFIE) screening test. The BEFIE test shows high specificity and sensitivity for absolute peripheral VFDs in neurologically impaired children (Koenraads et al., 2015). Additionally, the BEFIE test detects VFDs 4 years earlier than SAP (Portengen, Koenraads, Imhof, & Porro, 2020). However, limitations of the BEFIE test are the need of two assessors along with the inability to test the central VF and detect relative scotomas. PP circumvents these limitations and might be a suitable alternative to objectively test this patient group. Future studies may determine whether PP can map the VFs of children in an accurate, quick, and engaging way.

Conclusion

To conclude, we conducted the first in-depth comparison of three PP methods. All methods performed reasonably well in discerning simulated scotomas in healthy adults but gaze-contingent flicker pupil perimetry was superior in differentiating between (partially) present and absent stimuli whereas multifocal

pupil perimetry slightly better discerned present from partially present stimuli.

Keywords: perimetry, pupillometry, visual field, scotoma

Acknowledgments

Supported by the ODAS foundation (grant number 2017-03), the Rotterdamse Stichting Blindenbelangen (grant number B20170004), the F.P. Fischer Foundation (grant number 170511), and a grant from the Janivo Foundation (grant number 2017170). M.N. is supported by a grant from UitZicht (grant 2018-10, fund involved: Rotterdamse Stichting Blindenbelangen).

Commercial relationships: none.
Corresponding author: Brendan L. Portengen.
Email: b.l.portengen-2@umcutrecht.nl.
Address: Department of Ophthalmology, University Medical Center Utrecht, PO Box 85500, Room E 03.136, 3508 GA Utrecht, The Netherlands.

References

Allen, L. E., Slater, M. E., Proffitt, R. V., Quarton, E., & Pelah, A. (2012). A new perimeter using the preferential looking response to assess peripheral visual fields in young and developmentally delayed children. *Journal of AAPOS, 16*(3), 261–265.

Artes, P. H., Iwase, A., Ohno, Y., Kitazawa, Y., & Chauhan, B. C. (2002). Properties of perimetric threshold estimates from full threshold, SITA standard, and SITA fast strategies. *Investigative Ophthalmology and Visual Science, 43*(8), 2654–2659.

Binda, P., & Murray, S. O. (2015). Spatial attention increases the pupillary response to light changes. *Journal of Vision, 15*(2), 1.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436.

Buračas, G. T., & Boynton, G. M. (2002). Efficient design of event-related fMRI experiments using m-sequences. *NeuroImage, 16*(3 I), 801–813.

Carle, C. F., James, A. C., Kolic, M., Essex, R. W., & Maddess, T. (2015). Blue multifocal pupillographic objective perimetry in glaucoma. *Investigative Ophthalmology and Visual Science, 56*(11), 6394–6403.

Carle, C. F., James, A. C., Sabeti, F., Kolic, M., Essex, R. W., Shean, C., . . . Maddess, T. (2022). Clustered Volleys Stimulus Presentation for Multifocal Objective Perimetry. *Translational Vision Science and Technology, 11*(2), 5.

Castaldi, E., Pomè, A., Cicchini, G. M., Burr, D., & Binda, P. (2021). Pupil size automatically encodes numerosity. *Journal of Vision, 21*(9), 2302.

Chibel, R., Sher, I., Ben Ner, D., Mhajna, M. O., Achiron, A., Hajyahia, S., . . . Rotenstreich, Y. (2016). Chromatic Multifocal Pupillometer for Objective Perimetry and Diagnosis of Patients with Retinitis Pigmentosa. *Ophthalmology, 123*(9), 1898–1911.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, and Computers, 34*(4), 613–617.

Einhäuser, W. (2017). The pupil as marker of cognitive processes. In *Cognitive Science and Technology* (pp. 141–169). Singapore: Springer.

Gestefeld, B., Grillini, A., Marsman, J. B. C., & Cornelissen, F. W. (2020). Using natural viewing behavior to screen for and reconstruct visual field defects. *Journal of Vision, 20*(9), 1–16,

Harding, G. F. A., Spencer, E. L., Wild, J. M., Conway, M., & Bohn, R. L. (2002). Field-specific visual-evoked potentials: Identifying field defects in vigabatrin-treated children. *Neurology, 58*(8), 1261–1265.

Ho, Y.-L., Wong, S. S. Y., Carle, C. F., James, A. C., Kolic, M., Maddess, T., . . . Goh, X.-L. (2010). Multifocal Pupillographic Perimetry With White and Colored Stimuli. *Journal of Glaucoma, 20*(6), 336–343.

James, A. C., Kolic, M., Bedford, S. M., & Maddess, T. (2012). Stimulus parameters for multifocal pupillographic objective perimetry. *Journal of Glaucoma, 21*(9), 571–578.

Kardon, R. H. (1992). Pupil perimetry. In *Current Opinion in Ophthalmology 3*(5), 565–570.

Kardon, R. H., Kirkali, P. A., & Thompson, H. S. (1991). Automated Pupil Perimetry Pupil Field Mapping in Patients and Normal Subjects. *Ophthalmology, 98*(4), 485–496.

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What's new in Psychtoolbox-3? A free cross-platform toolkit for psychophysics with Matlab and GNU/Octave. In *Cognitive and Computational Psychophysics* (Vol. *36*). Retrieved from http://www.psychtoolbox.org.

Koenraads, Y., Braun, K. P. J., van der Linden, D. C. P., Imhof, S. M., & Porro, G. L. (2015). Perimetry in young and neurologically impaired children: the Behavioral Visual Field (BEFIE) Screening

Test revisited. *JAMA Ophthalmology, 133*(3), 319–325.

Maddess, T. (2014). Modeling the relative influence of fixation and sampling errors on retest variability in perimetry. *Graefe's Archive for Clinical and Experimental Ophthalmology, 252*(10), 1611–1619.

Maddess, T., Bedford, S. M., Goh, X. L., & James, A. C. (2009). Multifocal pupillographic visual field testing in glaucoma. *Clinical & Experimental Ophthalmology, 37*(7), 678–686.

Maddess, T., Essex, R. W., Kolic, M., Carle, C. F., & James, A. C. (2013). High- versus low-density multifocal pupillographic objective perimetry in glaucoma. *Clinical & Experimental Ophthalmology, 41*(2), 140–147.

Maeda, F., Kelbsch, C., Straßer, T., Skorkovská, K., Peters, T., Wilhelm, B., . . . Wilhelm, H. (2017). Chromatic pupillography in hemianopia patients with homonymous visual field defects. *Graefe's Archive for Clinical and Experimental Ophthalmology, 255*(9), 1837–1842.

Mathôt, S., & Van der Stigchel, S. (2015). New Light on the Mind's Eye: The Pupillary Light Response as Active Vision. *Current Directions in Psychological Science, 24*(5), 374–378.

Naber, M., Alvarez, G. A., & Nakayama, K. (2013). Tracking the allocation of attention using human pupillary oscillations. *Frontiers in Psychology, 4*, 919.

Naber, M., Roelofzen, C., Fracasso, A., Bergsma, D. P., van Genderen, M., Porro, G. L., . . . van der Schouw, Y. (2018). Gaze-Contingent Flicker Pupil Perimetry Detects Scotomas in Patients With Cerebral Visual Impairments or Glaucoma. *Frontiers in Neurology, 9*(July), 558.

Numata, T., Maddess, T., Matsumoto, C., Okuyama, S., Hashimoto, S., Nomoto, H., . . . Shimomura, Y. (2017). Exploring test-retest variability using high-resolution perimetry. *Translational Vision Science and Technology, 6*(5), 8.

Patel, D. E., Cumberland, P. M., Walters, B. C., Russell-Eggitt, I., & Rahi, J. S., & OPTIC Study Group. (2015). Study of Optimal Perimetric Testing in Children (OPTIC): Feasibility, Reliability and Repeatability of Perimetry in Children. *PLoS One, 10*(6), e0130895.

Pel, J. J. M., van Beijsterveld, M. C. M., Thepass, G., & van der Steen, J. (2013). Validity and Repeatability of Saccadic Response Times Across the Visual Field in Eye Movement Perimetry. *Translational Vision Science & Technology, 2*(7), 3.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

Piltz, J. R., & Starita, R. J. (1990). Test-retest variability in glaucomatous visual fields. In *American Journal of Ophthalmology 109*(1), 109–110.

Portengen, B. L., Koenraads, Y., Imhof, S. M., & Porro, G. L. (2020). Lessons Learned from 23 Years of Experience in Testing Visual Fields of Neurologically Impaired Children. *Neuro-Ophthalmology, 44*(6), 361–370.

Portengen, B. L., Roelofzen, C., Porro, G. L., Imhof, S. M., Fracasso, A., & Naber, M. (2021). Blind spot and visual field anisotropy detection with flicker pupil perimetry across brightness and task variations. *Vision Research, 178*(October 2020), 79–85.

Rajan, M. S., Bremner, F. D., & Riordan-Eva, P. (2002). Pupil perimetry in the diagnosis of functional visual field loss. *Journal of the Royal Society of Medicine, 95*(10), 498–500.

Rosli, Y., Carle, C. F., Ho, Y., James, A. C., Kolic, M., Rohan, E. M. F., . . . Maddess, T. (2018). Retinotopic effects of visual attention revealed by dichoptic multifocal pupillography. *Scientific Reports, 8*(1), 1–13.

Sabeti, F., James, A. C., Carle, C. F., Essex, R. W., Bell, A., & Maddess, T. (2017). Comparing multifocal pupillographic objective perimetry (mfPOP) and multifocal visual evoked potentials (mfVEP) in retinal diseases. *Scientific Reports, 7*, 45847.

Sabeti, F., James, A. C., Essex, R. W., & Maddess, T. (2013). Multifocal pupillography identifies retinal dysfunction in early age-related macular degeneration. *Graefe's Archive for Clinical and Experimental Ophthalmology, 251*(7), 1707–1716.

Sabeti, F., James, A. C., & Maddess, T. (2011). Spatial and temporal stimulus variants for multifocal pupillography of the central visual field. *Vision Research, 51*(2), 303–310.

Sabeti, F., Maddess, T., Essex, R. W., Saikal, A., James, A. C., & Carle, C. F. (2014). Multifocal pupillography in early age-related macular degeneration. *Optometry and Vision Science, 91*(8), 904–915.

Schmid, R., Luedtke, H., Wilhelm, B. J., & Wilhelm, H. (2005). Pupil campimetry in patients with visual field loss. *European Journal of Neurology, 12*(8), 602–608.

Skorkovská, K., Lüdtke, H., Wilhelm, H., & Wilhelm, B. (2009). Pupil campimetry in patients with retinitis pigmentosa and functional visual field loss. *Graefe's Archive for Clinical and Experimental Ophthalmology, 247*(6), 847–853.

Skorkovská, K., Wilhelm, H., Lüdtke, H., & Wilhelm, B. (2009). How sensitive is pupil campimetry in

hemifield loss? *Graefe's Archive for Clinical and Experimental Ophthalmology, 247*(7), 947–953.

Sutter, E. E. (1991). Fast m-transform. A fast computation of cross-correlations with binary m-sequences. *SIAM Journal on Computing, 20*(4), 686–694.

Sutter, E. E., & Tran, D. (1992). The field topography of ERG components in man-I. The photopic luminance response. *Vision Research, 32*(3), 433–446.

Tan, L., Kondo, M., Sato, M., Kondo, N., & Miyake, Y. (2001). Multifocal pupillary light response fields in normal subjects and patients with visual field defects. *Vision Research, 41*(8), 1073–1084.

Wall, M., Woodward, K. R., Doyle, C. K., & Artes, P. H. (2009). Repeatability of automated perimetry: A comparison between standard automated perimetry with stimulus size III and V, matrix, and motion perimetry. *Investigative Ophthalmology and Visual Science, 50*(2), 974–979.

Wilhelm, H., Neitzel, J., Wilhelm, B., Beuel, S., Lüdtke, H., Kretschmann, U., . . . Zrenner, E. (2000). Pupil Perimetry using M-Sequence Stimulation Technique. *Investigative Ophthalmology & Visual Science, 41*(5), 1229–1238.

Wong, A. M. F., & Sharpe, J. A. (2000). A comparison of tangent screen, goldmann, and humphrey perimetry in the detection and localization of occipital lesions. *Ophthalmology, 107*(3), 527–544.

Yoshitomi, T., Matsui, T., Tanakadate, A., & Ishikawa, S. (1999). Comparison of Threshold Visual Perimetry and Objective Pupil Perimetry in Clinical Patients. *Journal of Neuro-Ophthalmology, 19*(2), 89–99.