tvst

Article

Virtual Reality–Based and Conventional Visual Field Examination Comparison in Healthy and Glaucoma Patients

Jan Stapelfeldt[1,*], Şerife Seda Kucur[1,*], Nina Huber[1], René Höhn[2], and Raphael Sznitman[1]

[1] ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland
[2] Department of Ophthalmology, Bern University Hopsital, Bern, Switzerland

Correspondence: Şerife Seda Kucur, Murtenstrasse 50, 3008 Bern, Switzerland. e-mail: serife.kucur@unibe.ch

Received: February 26, 2021
Accepted: September 1, 2021
Published: October 6, 2021

Keywords: perimetry; visual field; virtual reality

Citation: Stapelfeldt J, Kucur ŞS, Huber N, Höhn R, Sznitman R. Virtual reality–based and conventional visual field examination comparison in healthy and glaucoma patients. Transl Vis Sci Technol. 2021;10(12):10, https://doi.org/10.1167/tvst.10.12.10

Purpose: Clinically evaluate the noninferiority of a custom virtual reality (VR) perimetry system when compared to a clinically and routinely used perimeter on both healthy subjects and glaucoma patients.

Methods: We use a custom-designed VR perimetry system tailored for visual field testing. The system uses Oculus Quest VR headset (Facebook Technologies, LLC, Bern, Switzerland), that includes a clicker for participant response feedback. A prospective, single center, study was conducted at the Department of Ophthalmology of the Bern University Hospital (Bern, Switzerland) for 12 months. Of the 114 participants recruited 70 subjects (36 healthy and 34 glaucoma patients with early to moderate visual field loss) were included in the study. Participants underwent perimetry tests on an Octopus 900 (Haag-Streit, Köniz, Switzerland) as well as on the custom VR perimeter. In both cases, standard dynamic strategy (DS) was used in conjunction with the G testing pattern. Collected visual fields (VFs) from both devices were then analyzed and compared.

Results: High mean defect (MD) correlations between the two systems (Spearman, $\rho \geq 0.75$) were obtained. The VR system was found to slightly underestimate VF defects in glaucoma subjects (1.4 dB). No significant bias was found with respect to eccentricity or subject age. On average, a similar number of stimuli presentations per VF was necessary when measuring glaucoma patients and healthy subjects.

Conclusions: This study demonstrates that a clinically used perimeter and the proposed VR perimetry system have comparable performances with respect to a number of perimetry parameters in healthy and glaucoma patients with early to moderate visual field loss.

Translational Relevance: This suggests that VR perimeters have the potential to assess VFs with high enough confidence, whereby alleviating challenges in current perimetry practices by providing a portable and more accessible visual field test.

Introduction

Glaucoma is one of the leading causes of blindness[1,2] and currently affects about 80 million people worldwide. Concurrently, the number of glaucoma patients is expected to rise toward 112 million by 2040.[3] Visual field (VF) testing is an integral part of glaucoma diagnosis and its follow-up monitoring[4] and standard automated perimetry (SAP) is the current gold standard VF testing approach.[5] Using perimeters, SAP quantitatively assesses an individual's visual function across the field of vision. More precisely, based on an interactive procedure involving participant responses to a sequence of light stimuli, SAP determines sensitivity thresholds (ST) at specific retinotopic locations, resulting in a measured VF.[6] Collected VFs can then be evaluated by comparing corresponding STs to an age-matched normative database to identify perceptual defects typical in glaucoma patients and other neuro-ophthalmic conditions.

While SAP is an essential tool in glaucoma diagnosis and monitoring, SAP has important limitations and weaknesses. Critically, SAP acquisitions, although fast

alternatives exist (e.g., 2–5 minutes),[7–10] are generally long-lasting with typical testing times around 6 to 8 minutes per eye,[6] during which the patient is required to maintain high levels of concentration.[11] Such long acquisition times, combined with the slightly forward leaning position of the participant has been reported to be uncomfortable and highly taxing for participants, leading to low-quality measurements,[11–14] as well as unwillingness in patients to participate in the follow-up examinations.[15,16] Furthermore, the management capacity in clinics is limited by the high-cost and large space requirements of standard perimeters and have already become overloaded with the increasing number of glaucoma patients.[17] As reported by Foot and MacEwen,[18] patients suffer from preventable loss due to the delays in the follow-up appointments, which highlights the incapacity of current eye care services.[16]

To ease these limitations, the use of portable alternatives has drawn strong attention. These have broadly been categorized in tablet or head-mounted device (HMD) systems. Tablet perimeters such as the VisualFields Easy,[19,20] Melbourne Rapid Fields (MRF),[21–27] or Eyecatcher[28,29] offer different solutions that use either gaze-based or touch-based feedback systems. These have shown to have fairly comparable test-retest variability with traditional perimeters under specific conditions.

HMD that used wired or wireless connections to a computer have also been used for standard VF testing. An early example was the tethered system presented by Chan et al.[30] The device showed good agreement with Humphrey Field Analyzer (HFA; Carl Zeiss, Jena, Germany) on 13 normal subjects and nine subjects with VF defects in terms of mean sensitivity (MS). A follow-up study using a virtual reality perimeter, Kasha Visual Field system, demonstrated that the system could achieve an equal sensitivity with neurosurgery patients.[31] Wroblewski et al.[32] introduced a HMD, VirtualEye, with two feedback options, namely manual (i.e., mouse click) and visual grasp (i.e., directing the gaze to the target). This system had a systematic bias of 4 to 6 dB and an average standard deviation of 5 dB compared to HFA while having fixation issues with visual grasp mode. The imo system (CREWT Medical Systems, Tokyo, Japan) implements monocular and binocular testing and showed high correlation with HFA on 20 glaucoma patients.[33] Due to the HMD's heaviness, the authors also propose a stationary stand to perform the test. Tsapakis et al.[34] introduced a VF testing system using virtual reality (VR) glasses for a smartphone and showed high correlation with HFA. Along this line, most recent HMD perimeters leverage the rapid advances in VR technology and are starting to be introduced as medical devices for clinical use. One such system is the Vivid Vision Perimeter (Vivid Vision, Inc., San Francisco, CA), with patient feedback based on gaze direction and was shown SAP consistent measurement variability.[35] Alternatively, the VisuALL (Olleyes Inc., Summit, NJ) VR perimeter showed moderate correlation with HFA (correlation coefficient $r = 0.5$ for normal, $r = 0.8$ for glaucoma subjects) while having longer examination times.[36]

In general, the aforementioned solutions could deliver portability in perimetry systems up to a limited extent. While some devices are not completely portable due to the device weight or cable requirement to connect to the computer, others either suffer from poor correlation with standard perimeters or induce longer examinations times. Furthermore, the comparison with the existing perimeters has usually been limited to individual STs or mean sensitivity values, which are device-dependent thus potentially misleading. In this paper, we introduce a novel VR perimeter and demonstrate its performance on a clinical population of both normal healthy subjects and glaucoma patients. We compare the system directly with a standard clinical perimeter, Octopus 900 (Haag-Streit). To perform a direct comparison between the two devices, we created a normative database for the proposed VR perimetry system. Mean defect (MD) values from both devices were contrasted to assess the noninferiority of the presented VR system.

Methods

We conducted a quantitative prospective randomized single-center study on 114 subjects with healthy and glaucomatous eyes in the Department of Ophthalmology, Bern University Hospital, from August 2019 to November 2020. The aim of the study was to evaluate the performance of a proposed VR perimetry system compared to a conventional perimeter.

All data captured throughout the study was stored either locally or using a secure web-based electronic data capturing tool. The data was anonymized for the data analysis. The study protocol with the identifier number 2018-01902 was approved by the Bernese Ethics Committee, Switzerland, and adhered to the tenets of the Declaration of Helsinki. Informed consent was obtained from all subjects in advance to the study procedure.

Subjects

Patients with healthy or glaucomatous eyes were recruited from the glaucoma clinic at the Department of Ophthalmology, Bern University Hospital. The general inclusion criteria were an age between

Table 1. Age, MD, and sLV Statistics of Subjects Included in the Study, Along with Mean, Standard Deviations (SD), and Max/Min Values

	#Patients	Age (Mean/SD [Min, Max])	MD (Mean/SD [Min, Max])	sLV (Mean/SD [Min, Max])
Healthy	36	60.67/10.64 [40.00, 77.00]	0.08/1.29 [–2.30, 4.00]	2.13/0.55 [1.10, 3.70]
Glaucoma	34	65.06/7.15 [53.00, 79.00]	5.69/2.53 [2.10, 10.30]]	6.26/2.63 [2.00, 11.80]
All	70	62.80/9.38 [40.00, 79.00]	2.80/3.44 [–2.30, 10.30]	4.14/2.79 [1.10, 11.80]

Figure 1. Oculus Quest VR headset with the Bluetooth connected clicker.

40 and 80 years, refractive error within ±5 diopter spherical equivalent, an astigmatism less than −3 diopter, a visual acuity of more than 0.3 logMar, a history of at least one perimetry examination and less than 25% false positive and negative errors for both examinations (VR or traditional system). Healthy subjects had mean defects (MD) of less than 2 dB; glaucomatous subjects were diagnosed with either primary open-angle, pseudoexfoliation, or primary angle-closure glaucoma, with early to moderate visual field loss (+2 dB < MD < +12 dB). Exclusion criteria were the inability to follow the procedure, insufficient knowledge of the project language (German or French), history of ocular diseases other than glaucoma or cataract, or any other visual pathway conditions that might affect visual field testing (e.g., pituitary lesions, demyelinating diseases). Patients having an history of epilepsy were also excluded.

In this study, 70 out of 114 subjects qualified for the study (36 healthy and 34 glaucomatous) meeting the inclusion criteria (i.e., false positive and negative errors less than 25% for each examination with the virtual reality system and the Octopus 900 perimeter).

Table 1 summarizes the statistics of the subjects included in the study.

Virtual Reality System

Our VR system uses the commercially available Oculus Quest VR headset (see Fig. 1) to simulate a standard visual field test. We implemented the same testing principles as in the Octopus 900 using Unity 2019.3.0 Alpha 5 (released on June 6, 2019). We used light stimuli of Goldmann size III for presentation on the VR screen which passes through the VR lens system to then be projected onto the retina of the subject. Subject feedback is provided with a remote clicker connected via Bluetooth.

Figure 2 shows the measured nonlinear relationship of the RGB inputs of the VR headset and the output luminance of the VR display. The standard background luminance level for perimetry of 10 cd/m^2 is equivalent to an RGB level of 94, enabling a theoretical dynamic range of 161 input levels, or 78 cd/m^2 luminance output range. Thus, to establish clinically relevant intensities for VF testing, the dB values are chosen so that the resolution is larger for increased dB values compared to lower dB values. Eq. 1 describes the used relation of luminance and decibel scale:

$$dB = 10 * \log_{10} \frac{L_{\max}}{L_{current} - L_{background}} \quad (1)$$

where L_{max} is the maximum luminance of the system (i.e., 88 cd/m^2), $L_{current}$ is the intensity of the displayed stimulus, and $L_{background}$ is 10 dB. To further correct spherical and cylindrical aberrations, custom designed lens holders were used to embed standard trial lenses in the VR system. The correction values were recomputed for the VR headset to establish the same lens correction as in Octopus 900.

To allow our system to produce mean defect values, we constructed a normative database using the data acquired from healthy subjects that were included in the study (See Table 1). Based on the Octopus 900's convention, we define a defect or total deviation (TD) at test location l, d_l as:

$$d_l = ST_l^{norm} - ST_l, \quad (2)$$

where ST_l^{norm} denotes the age-matched sensitivity threshold for the test location l and ST_l is the measured ST for the test location l. As an illustration, Figure 3 shows normative values with respect to the eccentricity of the test locations for subjects aged 55.

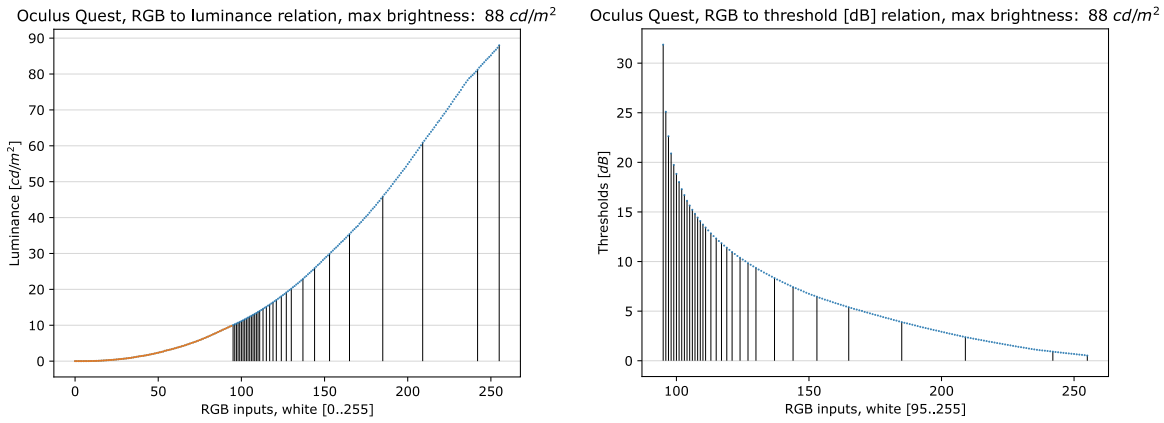

Figure 2. Screen calibration measurements of the VR headset shows relationship of RGB values to luminance and dB values, respectively. The vertical lines indicate the values that are used for the dynamic units of the testing strategy.

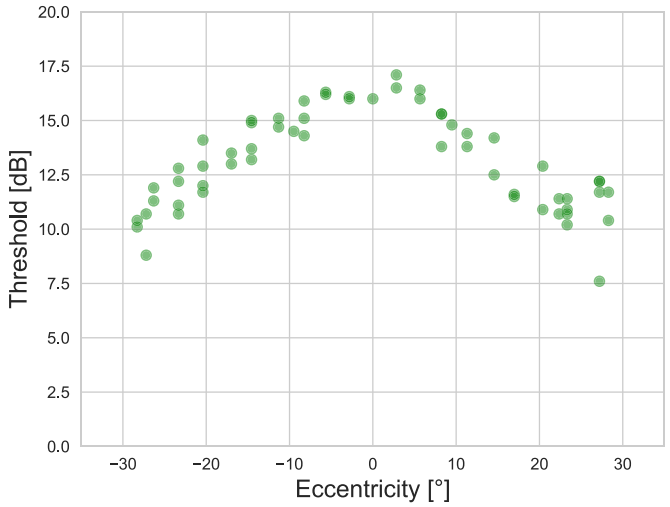

Figure 3. Example of sensitivity threshold values for 55-year-old subjects. Normative values decrease towards the periphery, showing the "hill of vision," as in other perimetry devices.[6]

Visual Field Acquisition

All subjects performed a visual field test using the custom designed VR system in addition to a visual field test using an Octopus900 (Haag-Streit). Both examinations took place the same day during the same session. VFs acquired with both devices used the dynamic strategy (DS),[37,38] which is routinely used for glaucoma patients in the Bern University Hospital Eye Clinic. The G program, that includes 59 test locations within 30°, was used for both systems as well.

For each subject, the order of two examinations was randomized to avoid fatigue-based biases. A randomly selected eye was chosen for the study in the case that both eyes met the defined inclusion criteria. Eye tracking option was turned off in Octopus 900 when testing the study eye, but the patient was monitored throughout the examination to ensure gaze fixation at the central target. As monitoring was not possible for the VR perimetry, we implemented blind spot control using the Heijl-Krakau method[39,40] to check the gaze fixation. A beeping sound was used for each test to maintain the subject's attention. There was (at least) a 5 minute break between Octopus 900 and VR device examinations.

Data Analysis/Statistics

To provide a direct comparison in MD differences between both systems, we will compute the correlations of corresponding VF using a Spearman ranked correlation coefficient (i.e., ρ, which takes values between –1 and 1, where 1/–1 indicates a strong relationship, while values closer to 0 suggests weak or no relationship).

Distribution of differences in MDs is also provided to show the existence of any *estimation bias*, whereby a distribution with a mean closer to 0 indicates an unbiased measurement. To investigate *performance dependency on the VF topography*, we also provide TD differences with respect to the eccentricity and spatial location. Additionally, we compute *Bland-Altman plots*[41] to determine the agreement between two measurement techniques. This allows for the identification systematic of bias as well as the limits of agreements that give the range of differences. In this case, it is defined to include 95% of the sample differences. If the range between the limits of agreement is clinically acceptable, then both systems can be used interchangeably. In addition, we computed distributions of examination times for both devices across different subpopulations. Similarly, to further analyze the differences between both device performance in terms of time, the

distributions of the number of presented stimuli are also given.

Illustrative examples are additionally provided to qualitatively assess VFs measured by the proposed VR system compared to those acquired by the Octopus 900. The error performance at the isolated defects is analyzed to further assess the acquisition ability of the VR system.

Lastly, we provide *comparison of reliability indices* (e.g., false positive/negative errors) of both devices.

Results

Figure 4 shows the correlations between the measured MDs for the Octopus 900 and the VR perimetry system. The Spearman's ranked correlation coefficients ρ is 0.77 (P value $<$ 0.00001), 0.50 ($P <$ 0.001), and 0.70 ($P <$ 0.0001) for all, healthy, and glaucoma patients, respectively.

Looking at the distribution of differences in MDs yielded means and standard deviations (SD) of 0.6 (SD = 2.3), –0.1 (SD = 2.2), and 1.4 (SD = 2.1) for all, healthy, and glaucoma subjects, respectively. The distributions are also presented in Figure 5 where no strong bias is apparent, but rather slight shifts towards underestimating MDs (compared to Octopus 900) in glaucoma subjects and slight overestimates are seen for healthy subjects. For all cases, however, the mean difference is less than 1.5 dB in both directions with a standard deviation of 2 dB. Figure 6 presents the MD estimation bias with respect to patient age. Accordingly, for healthy patients, VR perimetry shows slight overestimation for subjects aged less than 60 and underestimates subjects of age greater than 60. For glaucoma patients, VR perimetry tended to slightly underestimate across all age groups, while there is no statistically significant difference in the estimation bias with respect to the age ($P > 0.1$, Kruskal-Wallis test).

Figure 7 depicts the difference between total deviation measurements given by the two systems and distributions with respect to the eccentricity of the corresponding locations for both healthy and glaucomatous subject groups. Accordingly, no significant dependency is observed for either healthy or glaucomatous subjects (Kruskal-Wallis test, $P > 0.1$). For glaucomatous subjects, however, the standard deviation of differences slightly increases with increasing eccentricity. This can be further observed in Figure 8, which show TD differences (mean and standard deviation) with respect to the spatial location. For healthy patients, we observed homogeneous distribution of TD differences across the VF except a specific peripheral location where relatively higher underestimation by the VR system was observed. As for glaucoma cases, we observed higher TD bias at peripheral locations with no particular trend and relatively low variance in the inferior temporal region.

The Bland-Altman difference plots for all, normal, and glaucomatous subjects are given in Figure 9, with the middle line corresponding to the mean difference and the upper and lower lines corresponding to 95% limits of agreements (LoA).

Figure 10 compares the examination duration for Octopus 900 and VR perimetry for all, healthy, and glaucomatous patients separately. The median examination duration for the VR system is 6.57 minutes (CI: [6.41, 6.74]) while the median examination duration for Octopus 900 was 5.75 minutes (CI: [5.67, 5.96]) with $P < 0.0001$ (Kruskal-Wallis test). For healthy patients, there is a significant difference between VR median examination duration (6.70 minutes, CI: [6.50, 6.85]) and Octopus 900 examination duration (5.46 minutes, CI: [5.40, 5.67], $P < 0.0001$ (Kruskal-Wallis test)). In contrast, the difference is not statistically significant in glaucoma patients ($P > 0.1$, Kruskal-Wallis

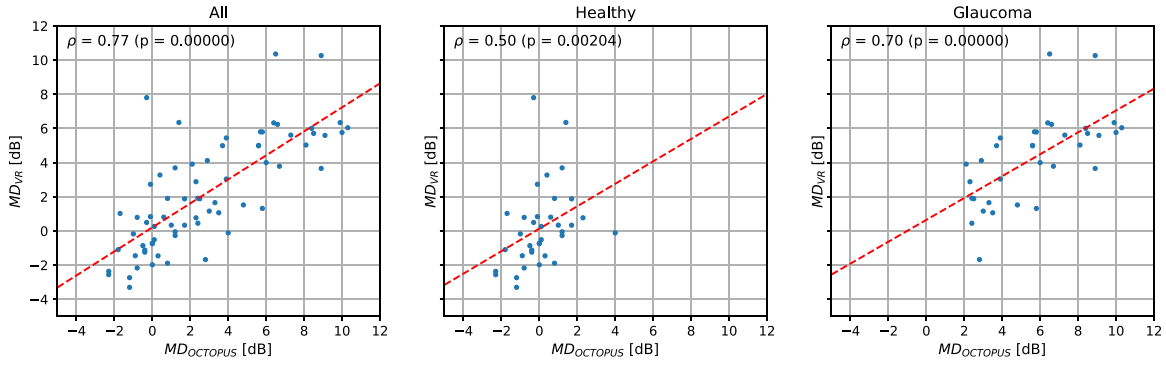

Figure 4. Mean defect correlations between all (*left*), healthy (*middle*), and glaucoma (*right*) subjects. Spearman's rank correlation coefficient ρ values with the corresponding P values are given on each plot. *Red dotted line* corresponds to best fit line.

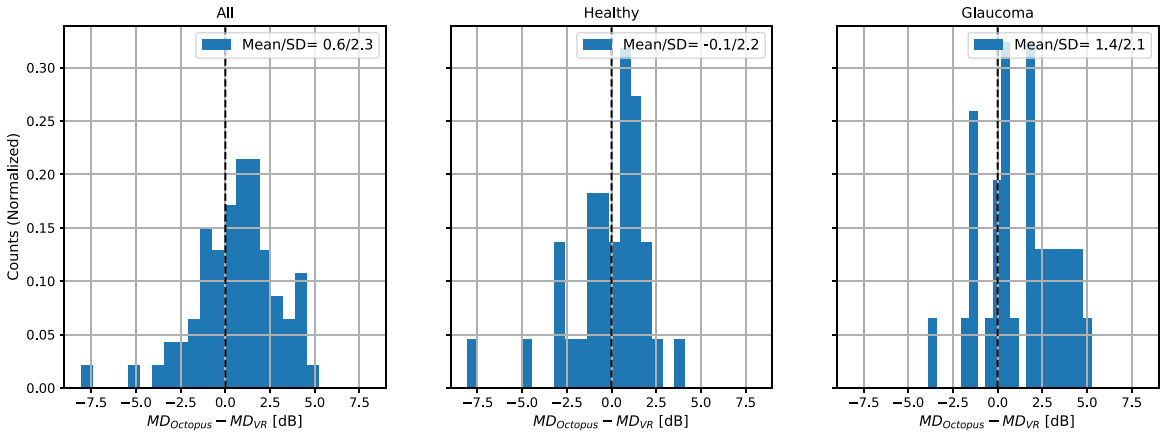

Figure 5. Estimation bias of MD measurements for all (*left*), healthy (*middle*), and glaucoma (*right*) patients. Mean and standard deviations (SD) are given in each plot.

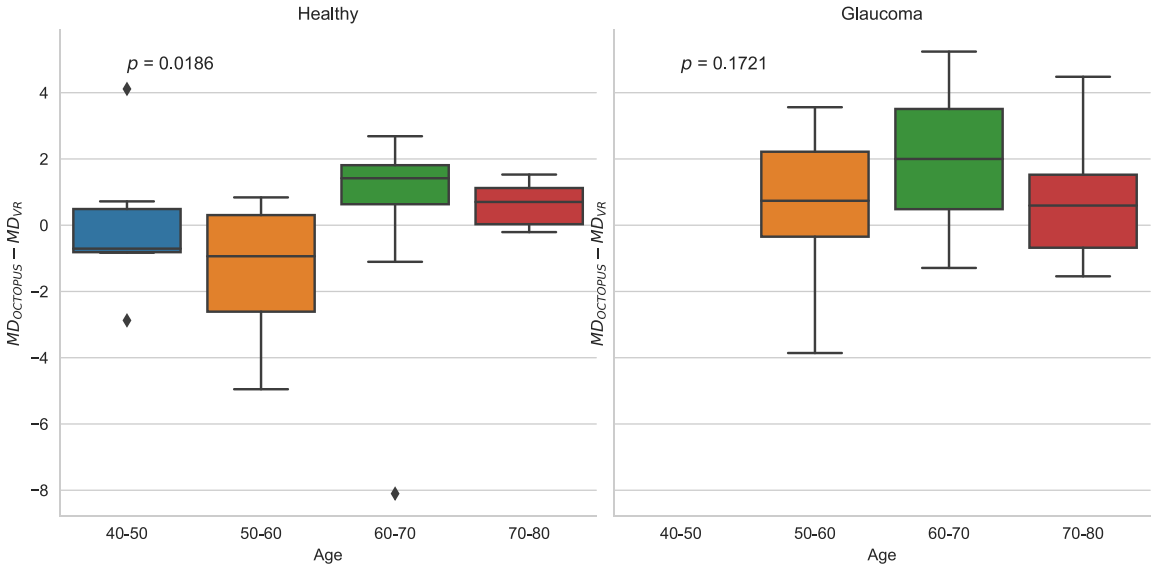

Figure 6. Estimation bias on MD estimation with respect to the age of the patient for healthy (*left*) and glaucoma (*right*) patients. *P* values are provided on each plot (Kruskal-Wallis test).

test) between VR (6.39 minutes, CI: [6.20, 6.73]) and Octopus 900 (6.27 minutes, CI: [5.97–6.65]) devices.

Figure 11 shows the distribution of the number of presented stimuli (excluding all the catch trials) by Octopus 900 and by the VR perimetry for all, healthy, and glaucoma subjects. The median number of stimuli presented by the VR system is 136.50 (CI: [133.50, 139.50]) while the median number of stimuli presented by Octopus 900 is 129.50 (CI: [126.00, 132.00]). For healthy patients, the number of stimuli presented by the VR system (139.00, CI: [134.00, 142.00]) is larger than that of Octopus 900 (125.00, CI: [124.00–126.50]); $P < 0.0.0001$, Kruskal-Wallis test). For glaucoma patients, VR system (135.00, CI: [129.00–139.00]) presented slightly fewer stimuli than Octopus 900 (137.00, CI: [132.00–150.00]; $P > 0.1$, Kruskal-Wallis test).

Qualitative examples for healthy (see Fig. 12) and glaucomatous (see Fig. 13) VFs are given. The total deviation values are scaled to allow comparison between Octopus 900 and VR perimetry acquisitions. In addition, we provide the performance of VR perimetry when measuring isolated defects in Figure 14. Here we see the differences between total deviation values of both systems at individual locations with respect to gradient measure, Δ_l, computed as described in the literature.[42,43] A high Δ_l means the location l is within a nonhomogeneous region and potentially a deep localized defect, which is important but more difficult to exactly capture in a visual field testing. The median error on relatively low Δ_l values, $\Delta_l < 15$ dB, is less than 5 dB for each group of subjects. As expected, there are few occurrences of isolated nonhomogeneous

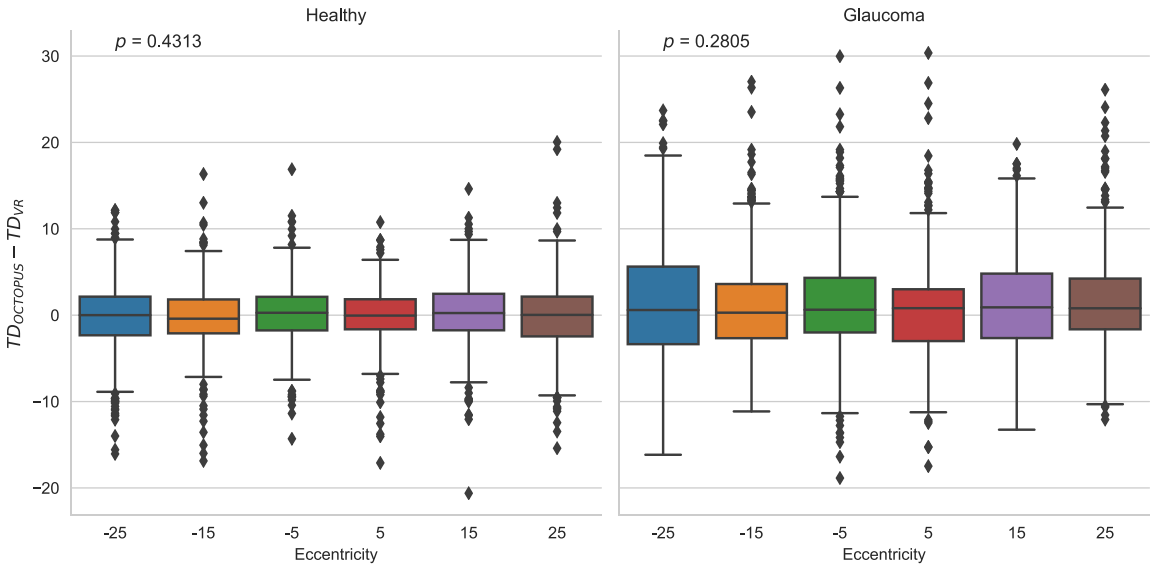

Figure 7. Estimation bias on total deviation with respect to the eccentricity of the patient for healthy (*left*) and glaucoma (*right*) subjects. *P* values are provided on each plot (Kruskal-Wallis test).

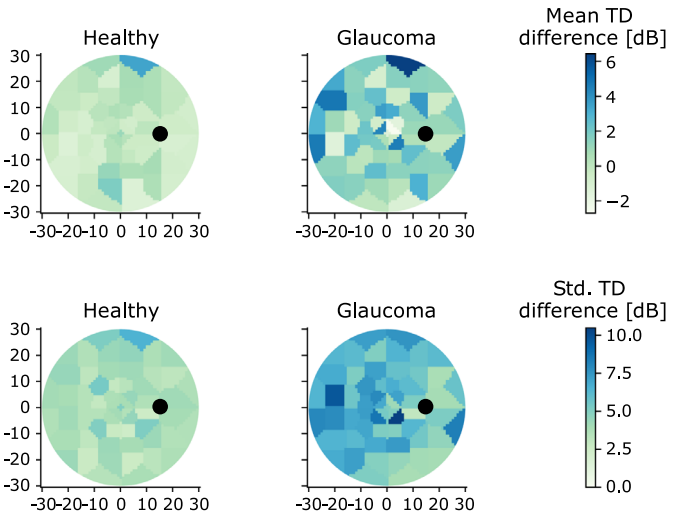

Figure 8. Mean (*top row*) and standard deviation (*bottom row*) of TD differences with respect to the spatial location for healthy (*left*) and glaucoma (*right*) subjects.

regions in healthy VFs. For relatively high Δ_l values (i.e., $\Delta_l \approx 20$) dB, the median error is less than 3 dB with higher standard deviation. As for glaucomatous VFs, the median error is less than 8 dB for $\Delta_l \in [17-27$ dB] and is within [10−15 dB] for larger Δ_l.

Regarding reliability comparison, we provide the distributions of false positive (FPR) and false negative response rates (FNR) for each device including all (114) patients (without excluding patients based on high FPR/FNR). The comparison given in Figure 15 and Table 2 show the distributions and the statistics for healthy and glaucoma subjects separately to assess if there is bias with respect to the patient's VF state. FPR of the proposed VR system is very similar to that of Octopus 900 for both healthy and glaucoma groups, even slightly lower in the VR (statistically significant for the healthy group, Mann-Whitney *U* test, $P < 0.05$; statistically not significant for the glaucoma group, $P > 0.1$). FNR, however, tends to be slightly higher in the VR system than in the Octopus 900 for both groups (statistically significant for both healthy and glaucoma groups, Mann-Whitney *U* test, $P < 0.05$).

Discussion

Mean defect comparison of the Octopus 900 and the VR perimetry system using Spearman's ranked correlation showed good agreement between both device measurements with ρ being larger than 0.7 for all and glaucomatous subjects. The lower ρ value obtained for the healthy subjects can be explained by the low variance of the corresponding data values (within the range of -2.5 dB $\leq$ MD $\leq$ 2.5 dB), which reduces the correlation coefficient (Constraining data to a small range is known to reduce the correlation coefficient, although the correlation on the whole range is high). Therefore, the correlation plots including all patients or glaucoma patients provide more meaningful information with regards to the correlation between the two measurements and demonstrated that the VR MDs are well correlated with Octopus 900 MDs.

The estimation bias for MD measurements provided in Figure 5 suggests almost no bias for healthy subjects and a slight underestimation for glaucoma patients.

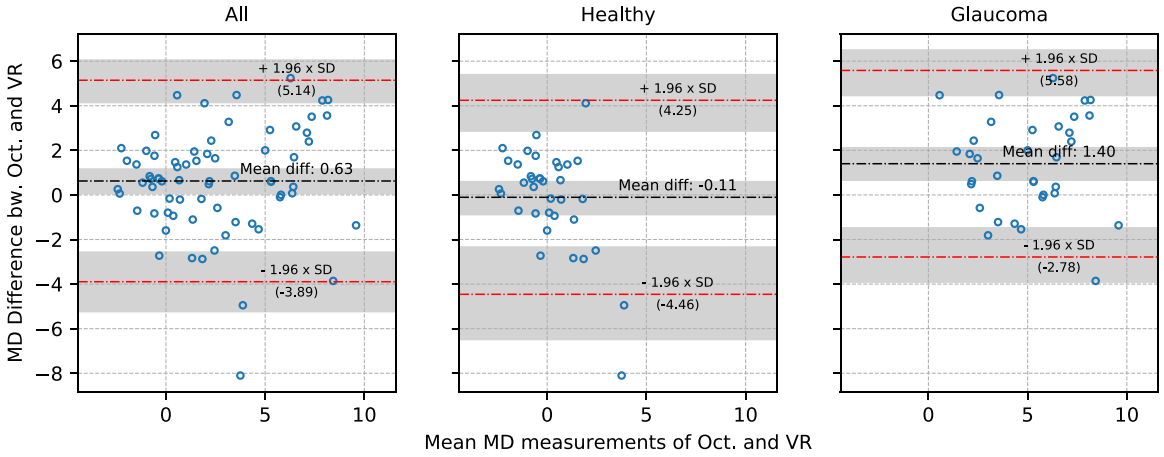

Figure 9. Bland-Altman agreement graphs between Octopus 900 and VR device MD measurements. The *black dotted line* corresponds to the mean difference and *red dotted lines* correspond to 95% limits of agreements (mean ± 1.96).

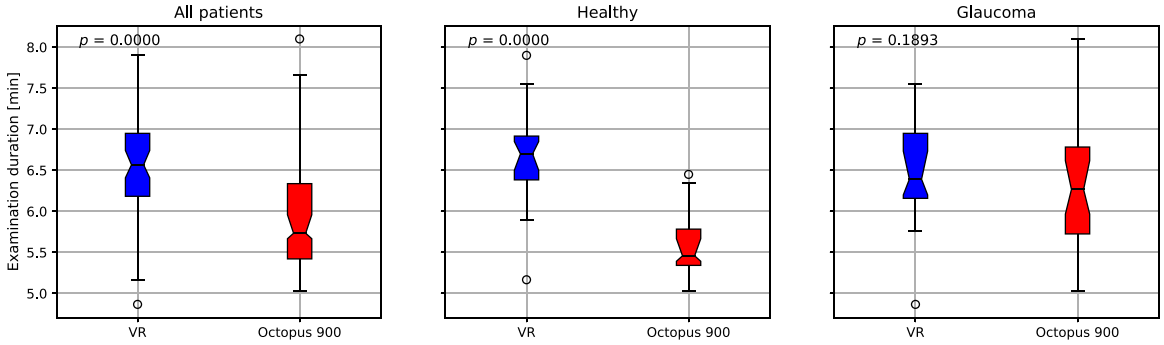

Figure 10. Distributions of examination duration for VR perimetry and Octopus 900 presented for all (*left*), healthy (*middle*), and glaucoma patients (*right*). For each subplot, *P* values are provided (Kruskal-Wallis test).

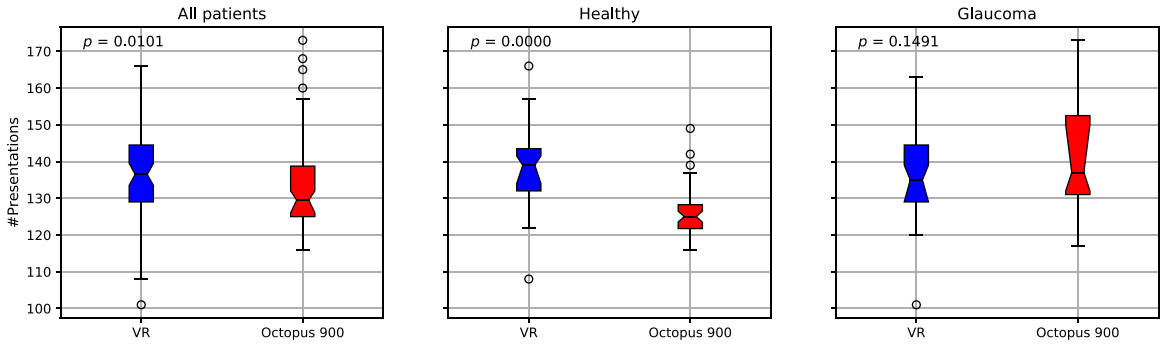

Figure 11. Distributions of number of presented stimuli for VR perimetry and Octopus 900 for all (*left*), healthy (*middle*), and glaucoma patients (*right*). For each subplot, *P* values are provided (Kruskal-Wallis test).

However, considering all patients, the bias is negligible with a mean difference of less than 1 dB. Interestingly, the standard deviations of estimation differences are almost the same for normal and glaucoma subjects. There was no dependence on particular age groups in glaucoma patients (Kruskal-Wallis test, $P > 0.1$) whereas, in healthy patients, we observed that there is slight overestimation of MDs for age group below 60 and underestimation for the group above 60 (Fig. 6). This can be partially explained by the fact that with the increasing age subjects have relatively reduced sensitivities, which our proposed system may not well acquire due to its luminance limits. The overestimation of MD by the VR system for relatively younger subjects

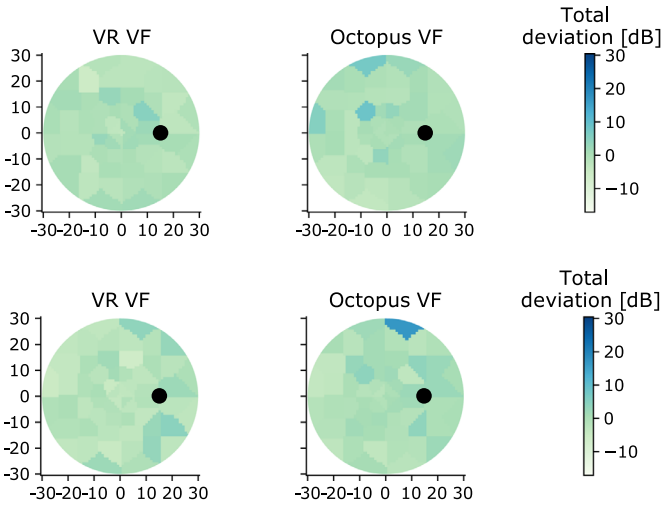

Figure 12. Two healthy VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). The *black circle* corresponds to the blind spot. *Blue colors* reflect higher deviations, that is, deeper defects.

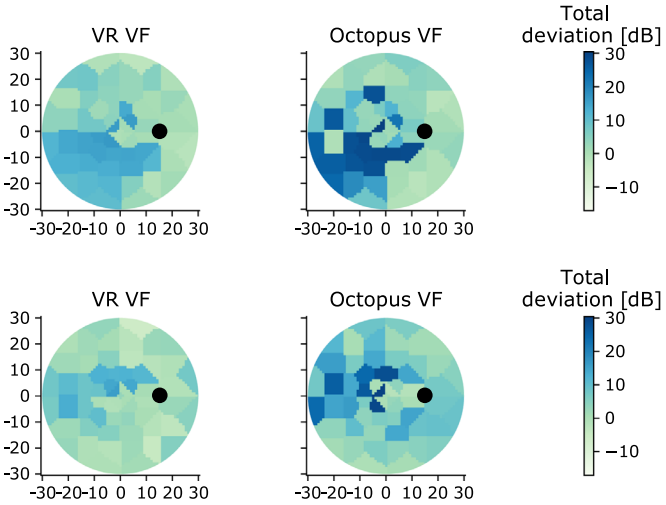

Figure 13. Two glaucomatous VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). The *black circle* corresponds to the blind spot. *Blue colors* reflect higher deviations, that is, deeper defects.

should, however, be further investigated and verified with more significant amount of data.

We also investigated the dependency of performance with respect to eccentricity and spatial location as given in Figure 7 and Figure 8. Accordingly, no significant bias was observed with respect to the eccentricity as shown in Figure 7 (Kruskal-Wallis test, $P > 0.1$). As for the distribution of the TD differences across the VF, we did not observe a clear trend of dependency in either group; yet higher bias in the peripheral region can be observed, especially in the glaucoma group (Fig. 8). Interestingly, for healthy subjects, there is only one location in the superior hemisphere where VR remarkably underestimated with high variance. Early visual field loss is known to start developing from the peripheral region and mostly in the superior hemifield,[44,45] which can explain high bias (i.e., underestimation by VR) and variance on those regions due to both high response variability and limited luminance of the proposed system. Especially, the variance distribution appears to follow the expansion of glaucomatous loss that affects has a lesser effect on the inferior temporal region[44,45] where the variance of TD difference was also observed to be low.

Alternatively, the Bland-Altman plots in Figure 9 illustrate the MD differences and stratify the MD differences into the mean MD measurements. The limits of agreements obtained for each group are found to be comparable and even better than those reported in similar comparative studies in the literature.[23,46] This indicates that the differences between MD measurements are within acceptable error range.

While the examination duration was found to be longer ($\approx$ 1.3 minutes longer) for healthy patients (Kruskal-Wallis test, $P < 0.0001$) and slightly longer for glaucoma patients (no statistical significance, Kruskal-Wallis test, $P > 0.1$), we mainly attribute this difference to the blind spot catch trials implemented in the VR system, as well as to some implementation differences of perimetry strategies between the two devices. When excluding the number of catch trials (see Fig. 11), we observed that healthy patients required more stimuli while glaucoma patients required slightly less than Octopus 900. The relatively higher number of stimuli required by healthy patients is potentially due to the smaller step sizes of our DS implementation in the VR device within the dimmer intensity range. Conversely, the smaller number of stimuli required by glaucoma patients is due to the limited intensity range: since the VR cannot present stimuli beyond the maximum intensity of the VR headset, which is less than Octopus 900, this naturally decreases the number of stimuli presented to glaucoma patients, especially at locations with deeper defects.

Qualitative examples for healthy VFs acquired by VR perimetry reflect overall coherence with those acquired by Octopus 900 while some localized deep defects in the VF by Octopus 900 may not be observed in VR perimetry acquired VFs (see the bottom example in Fig. 12). In some healthy cases, we observe localized defects in the VR-acquired VFs as well, which did not appear in their Octopus 900-acquired reciprocals (see Appendix). This may be due to implementation differences in DS, as well as to the occasional

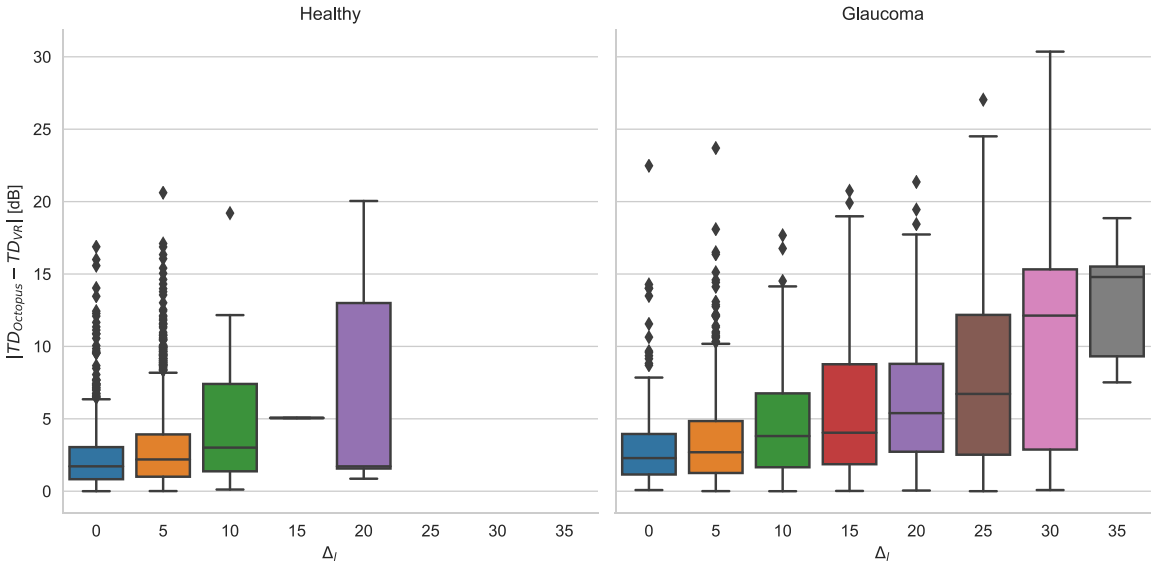

Figure 14. The differences between individual total deviation values with respect to the gradient measure Δ_I given for healthy and glaucoma subjects separately. A high Δ_I value indicates that the corresponding locations are inside a heterogeneous region, which is more difficult to accurately measure.

subject-specific reasons (e.g., novel experience with the VR system). The existence of isolated defects in the VR-acquired healthy VFs should be further investigated with a test-retest study to better conclude on the proposed system's behavior. As for glaucoma examples, we observed that VR perimetry could detect the spread of defect regions correctly, but the defect magnitudes are slightly lower than those given by Octopus 900 (see Fig. 13 and Appendix). This bias is coherent with Figure 5, where MDs were estimated lower by VR than by Octopus 900 for glaucoma patients. To better interpret the qualitative results, Figure 14 presents the individual errors (i.e., differences between both systems) with respect to the gradient measure. Here, the VR system had larger errors measuring the defects at locations with high gradient measure (i.e., less homogeneous region as also seen in the qualitative results). Yet, the error remains less than 8 dB for the gradient measures smaller than 25 dB and only gets very high for extremely deep and isolated defects (i.e., $\Delta_I > 27$ dB). These high errors can be attributed to the technical limits of the VR headset having a limited range of luminance which restricts the measurement of deep defects. Nonetheless, VR perimetry succeeds in capturing defect patterns, although occasionally fails to identify isolated defects.

In terms of reliability indices, Figure 15 shows that the proposed VR system has similar reliability behavior to the Octopus 900. We note that the FNR tends to be slightly higher in healthy and glaucoma groups when using our system. Such high FNRs may be due to the limited luminance in our system which cannot produce sufficiently higher intensity stimuli that glaucoma patients with relatively higher visual field defects need. However, FNRs generally remain under a reasonable value (FNR $\leq$ 0.2) for each method and for each group. Moreover, FNR as an indicator of patience performance has been criticized and has been shown to correlate with patient's visual field loss more than the patient's reliability.[47–49] This is also seen in Figure 15: FPR distributions of each device minimally change with respect to the patient group, whereas the distributions of FNR change strongly based on the patient group. Moreover, FNRs are higher for glaucoma groups supporting the claim that FNR may indicate more the VF state of the patient rather than their performance.

There are several drawbacks of the presented work. First, the proposed VR system has a limited luminance range that does not allow advanced losses to be measured as accurately as in conventional perimeters. Another shortfall is the need for the calibration of the VR screen luminosity, which may be affected by various internal and external factors (i.e., device temperature, battery level, etc.). This may induce additional noise into the measurements. Moreover, the current clinical study did not investigate the test-retest variability as well as the diagnostic utility of the proposed system, which are important metrics in assessing the performance of a perimeter. Given these limitations,

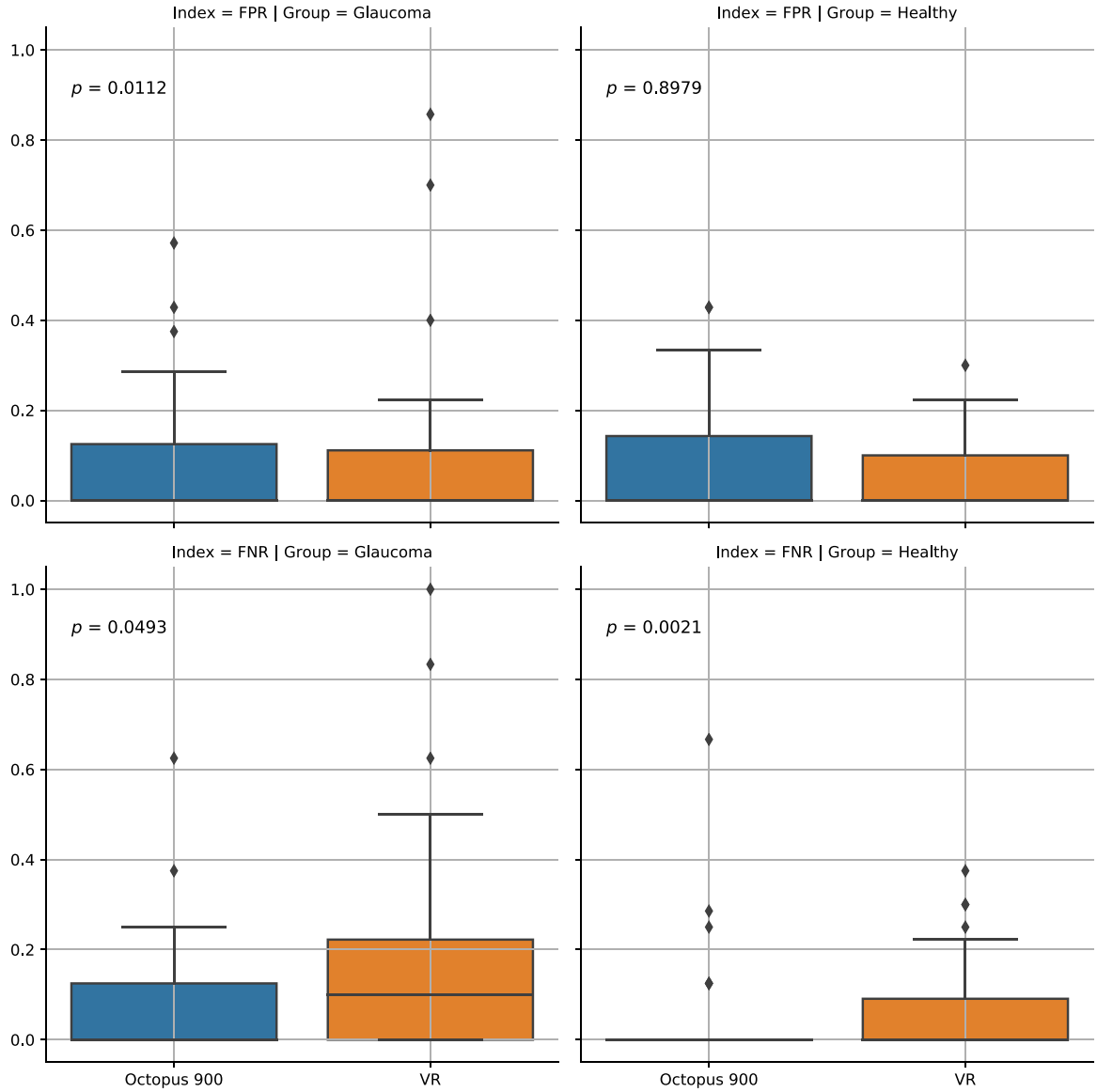

Figure 15. Distributions of the reliability indices, namely false positive (*left*)/negative (*right*) response rates of each device. For each subplot, *P* values are provided (Mann-Whitney *U* test).

our future work will focus on overcoming the VR perimetry technical limits to standardize the measurements, generating a normative database based on a larger normal healthy population to assess its diagnostic performance and designing a test-retest clinical study to show the test repeatability performance.

Table 2. Summary of Catch Trial False Positive Rates (FPR) and False Negative Rates (FNR) for Healthy and Glaucomatous subjects. Shown are the Median, Mean and Standard Deviation (SD) for Each Type of Catch Trial and Subject Group

		FPR (Median/Mean/SD)	FNR (Median/Mean/SD)
Healthy	Octopus 900	0.00/0.08/0.12	0.00/0.03/0.11
	VR	0.00/0.04/0.07	0.00/0.05/0.09
Glaucoma	Octopus 900	0.00/0.07/0.12	0.00/0.07/0.11
	VR	0.00/0.08/0.15	0.10/0.16/0.21

Conclusions

This study presents clinical results comparing the performances of a novel virtual reality–based perimetry system with a conventional stationary perimeter, Octopus 900. The proposed VR perimetry system simulates a standard visual field testing by presenting a sequential light stimulus to the patient who gives feedback with an associated clicker. In this study, we evaluated the proposed system on 70 subjects, including normal healthy subjects and glaucoma patients. The VFs acquired by the VR perimetry showed high correlation ($\rho > 0.7$) with those given by the Octopus 900. Qualitatively, VR perimetry could correctly identify the impaired areas in a VF while deep isolated defects were occasionally underestimated. With its portability and comfort, VR perimetry has great potential to become an alternative to conventional perimeters in the future.

Acknowledgments

Supported by funding from BRDIGE Proof of Concept grant (Project no: 40B1-0_184179).

Disclosure: **J. Stapelfeldt,** None; **Ş.S. Kucur,** None; **N. Huber,** None; **R. Höhn,** None; **R. Sznitman,** None

* JS and ŞSK contributed equally to this work.

References

1. Kingman S. Glaucoma is second leading cause of blindness globally. *Bull World Health Organ*. 2004;82:887–888.
2. Resnikoff S, Pascolini D, Etya'ale D, Kocur I, Pararajasegaram R, Pokharel GP, Mariotti SP. Global data on visual impairment in the year 2002. *Bull World Health Organ*. 2004;82:844–851.
3. Tham Y-C, Li X, Wong TY, Quigley HA, Aung T, Cheng C-Y. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121:2081–2090.
4. European Glaucoma Society, *Terminology and Guidelines for Glaucoma*. 4th ed. Savona, Italy: Publicomm; 2014.
5. Alencar LM, Medeiros FA. The role of standard automated perimetry and newer functional methods for glaucoma diagnosis and follow-up. *Indian J Ophthalmol*. 2011;59(Suppl):S53–S58.
6. Racette C, Lyne and Fischer, Monika and Bebie, Hans and Holló, Gábor and Johnson, Chris and Matsumoto, *Visual Field Digest*. 2016.
7. Morales J, Weitzman ML, González de la Rosa M. Comparison between tendency-oriented perimetry (TOP) and Octopus threshold perimetry. *Ophthalmology*. 2000;107:134–142.
8. Bengtsson B, Olsson J, Heijl A, Rootzén H. A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmol Scand*. 2009;75:368–375.
9. Bengtsson B, Heijl A. SITA Fast, a new rapid perimetric threshold test. Description of methods and evaluation in patients with manifest and suspect glaucoma. *Acta Ophthalmol Scand*. 1998;76:431–437.
10. Heijl A, Patella VM, Chong LX, et al. A new SITA perimetric threshold testing algorithm: construction and a multicenter clinical study*Am J Ophthalmol*. 2019;198:154–165.
11. Marra G, Flammer J. The learning and fatigue effect in automated perimetry. *Graefes Arch Clin Exp Ophthalmol*. 1991;229(6):501–504.
12. Johnson CA, Adams CW, Lewis RA. Fatigue effects in automated perimetry. *Appl Opt*. 1988;27(6):1030–1037.
13. Heijl A, Drance SM. Changes in differential threshold in patients with glaucoma during prolonged perimetry. *Br J Ophthalmol*. 1983;67:512–516.
14. Gonzalez de la Rosa M, Pareja A. Influence of the "fatigue effect" on the mean deviation measurement in perimetry. *Eur J Ophthalmol*. 1997;7(1):29–34.
15. Glen FC, Baker H, Crabb DP. A qualitative investigation into patients' views on visual field testing for glaucoma monitoring. *BMJ Open*. 2014;4(1), doi:10.1136/bmjopen-2013-003996.
16. Prior M, Francis JJ, Azuara-Blanco A, Anand N, Burr JM. Why do people present late with advanced glaucoma? A qualitative interview study. *Br J Ophthalmol*. 2013;97(12):1574–1578.
17. Broadway DC, Tibbenham K. Tackling the NHS glaucoma clinic backlog issue. *Eye*. 2019;33(11):1715–1721.
18. Foot B, MacEwen C. Surveillance of sight loss due to delay in ophthalmic treatment or review: frequency, cause and outcome. *Eye*. 2017;31(5):771–775.
19. Santos A, Morabe E. "VisualFields Easy": an iPad application as a simple tool for detecting visual field defects. *Phillip J Ophthalmol*. 2016;41:22–26.
20. Johnson CA, Thapa S, George Kong YX, Robin AL. Performance of an iPad application to detect moderate and advanced visual field loss in nepal. *Am J Ophthalmol*. 2017;182:147–154.
21. Vingrys AJ, Healey JK, Liew S, et al. Validation of a tablet as a tangent perimeter. *Trans Vis Sci Tech*. 2016;5:3–3.
22. Kong YXG, He M, Crowston JG, Vingrys AJ. A comparison of perimetric results from a tablet

perimeter and Humphrey Field Analyzer in glaucoma patients. *Trans Vis Sci Tech*. 2016;5:2.
23. Kong YXG, He M, Crowston JG, Vingrys AJ. A comparison of perimetric results from a tablet perimeter and Humphrey Field Analyzer in glaucoma patients. *Trans Vis Sci Tech*. 2016;5(6):2–2.
24. Anderson AJ, Bedggood PA, George Kong YX, Martin KR, Vingrys AJ. Can home monitoring allow earlier detection of rapid visual field progression in glaucoma? *Ophthalmology*. 2017;124(12):1735–1742.
25. Nesaratnam N, Thomas PB, Kirollos R, Vingrys AJ, Kong GY, Martin KR. Tablets at the bedside-iPad-based visual field test used in the diagnosis of intrasellar haemangiopericytoma: a case report. *BMC Ophthalmol*. 2017;17(1):53.
26. Prea SM, Kong YXG, Mehta A, et al. Six-month longitudinal comparison of a portable tablet perimeter with the Humphrey Field Analyzer. *Am J Ophthalmol*. 2018;190:9–16.
27. Kumar H, Thulasidas M. Comparison of perimetric outcomes from Melbourne rapid fields tablet perimeter software and Humphrey Field Analyzer in glaucoma patients. *J Ophthalmol*. 2020;2020, doi:10.1155/2020/8384509.
28. Jones PR, Smith ND, Bi W, Crabb DP. Portable perimetry using eyetracking on a tablet computer—a feasibility assessment. *Trans Vis Sci Tech*. 2019;8(1):17–17.
29. Jones PR, Lindfield D, Crabb DP. Using an open-source tablet perimeter (eyecatcher) as a rapid triage measure for glaucoma clinic waiting areas. *Br J Ophthalmol*. 2020, doi:10.1364/VSIA.1999.MB1.
30. Chan AD, Eizenman M, Flanagan JG, Trope GE. Head-mounted perimetry. In: *Vision Science and Its Applications*, p. MB1, Optical Society of America, 1999.
31. Hollander DA, Volpe NJ, Moster ML, et al. Use of a portable head mounted perimetry system to assess bedside visual fields. *Br J Ophthalmol*. 2000;84(10):1185–1190.
32. Wroblewski D, Francis BA, Sadun A, Vakili G, Chopra V. Testin of visual field with virtual reality goggles in manual and visual grasp modes. *Biomed Red Int*. 2015;2014:206082, doi:10.1155/2014/206082. Epub 23 June 2014.
33. Matsumoto C, Yamao S, Nomoto H, et al. Visual field testing with head-mounted perimeter 'imo'. *PLoS One*. 2016;11(8):1–12.
34. Tsapakis S, Papaconstantinou D, Diagourtas A, et al. Visual field examination method using virtual reality glasses compared with the Humphrey perimeter. *Clin Ophthalmol*. 2017;11:1431–1443.
35. Matsumoto JA, Deiner M, Nguyen A, et al. Measurement reproducibility using vivid vision perimetry: a virtual reality-based mobile platform. *Invest Ophthalmol Vis Sci*. 2020;61(7):4800.
36. Razeghinejad R, Gonzalez-Garcia A, Myers JS, Katz LJ. Preliminary report compared with standard automated perimetry. *J Glaucoma*. 2021;30(1):17–23.
37. Weber J. [A new strategy for automated static perimetry]. *Fortschr Ophthalmol*. 1990;87(1):37–40.
38. Weber J, Klimaschka T. Test time and efficiency of the dynamic strategy in glaucoma perimetry. *Ger J Ophthalmol*. 1995;4:25–31.
39. Heijl A, Krakau CET. An automatic static perimeter, design and pilot study. *Acta Ophthalmol (Copenh)*. 1975;53(3):293–310.
40. Heijl A, Krakau CE. A note of fixation during perimetry. *Acta Ophthalmol (Copenh)*. 1977;55:854–861.
41. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *J R Stat Soc Series D (The Statistician)*. 1983;32(3):307–317, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.384.3483&rep=rep1&type=pdf.
42. Chong LX, McKendrick AM, Ganeshrao SB, Turpin A. Customized, automated stimulus location choice for assessment of visual field defects. *Invest Opthalmol Vis Sci*. 2014;55:3265.
43. Kucur SS, Sznitman R. Sequentially optimized reconstruction strategy: a meta-strategy for perimetry testing. *PLoS One*. 2017;12(10):e0185049.
44. Kim JM, Kyung H, Shim SH, Azarbod P, Caprioli J. Location of initial visual field defects in glaucoma and their modes of deterioration. *Invest Ophthalmol Vis Sci*. 2015;56:7956–7962.
45. Yousefi S, Sakai H, Murata H, et al. Asymmetric patterns of visual field defect in primary open-angle and primary angle-closure glaucoma. *Invest Ophthalmol Vis Sci*. 2018;59:1279–1287.
46. Ichhpujani P, Lo DC, Cvintal V, et al. Flicker defined form, standard perimetry and Heidelberg retinal tomography: structure-function relationships. *Can J Ophthalmol*. 2015;50(4):290–296.
47. Katz J, Sommer A, Witt K. Reliability of visual field results over repeated testing. *Ophthalmology*. 1991;98(1):70–75.
48. Bengtsson B. Reliability of computerized perimetric threshold tests as assessed by reliability indices and threshold reproducibility in patients with suspect and manifest glaucoma. *Acta Ophthalmol Scand*. 2000;78(5):519–522.
49. Bengtsson B, Heijl A. False-negative responses in glaucoma perimetry: indicators of patient performance or test reliability?. *Invest Ophthalmol Vis Sci*. 2000;41(8):2201–2204.

Appendix: Qualitative Examples

In this appendix, we present qualitative comparisons of two acquired VFs for each subject included into the study.

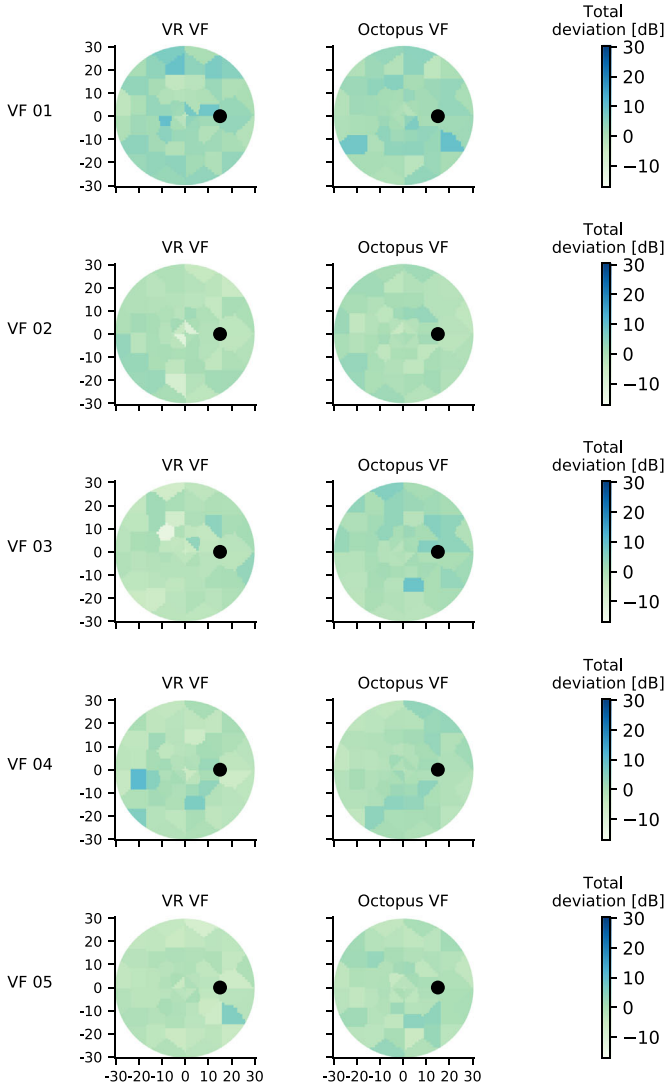

Figure 16. Five healthy VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is, deeper defects.

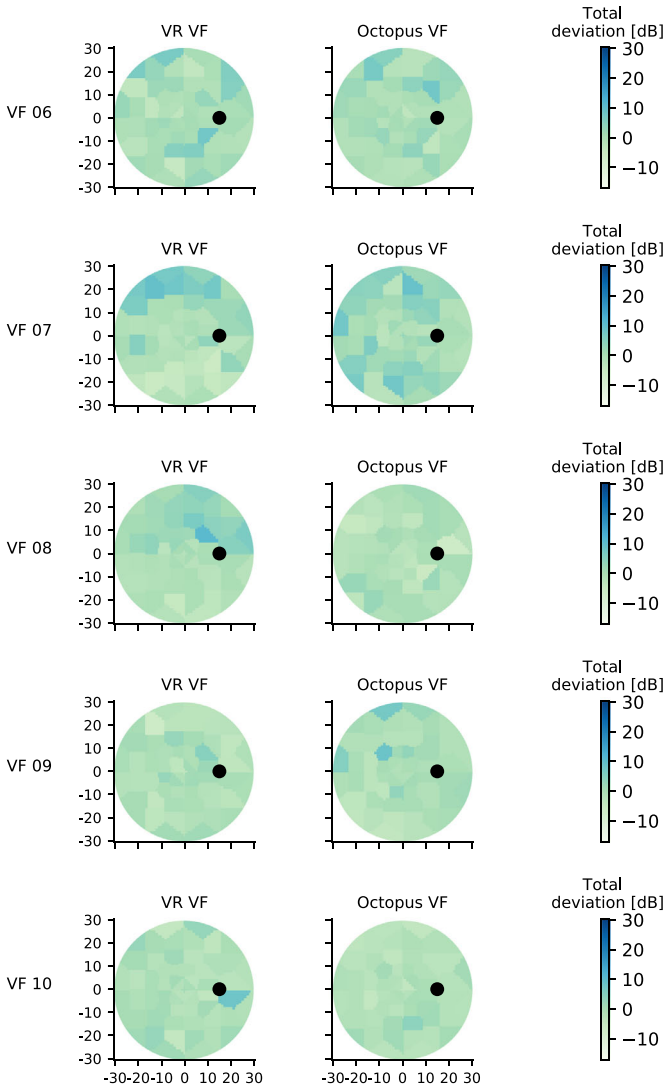

Figure 17. Five healthy VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is, deeper defects.

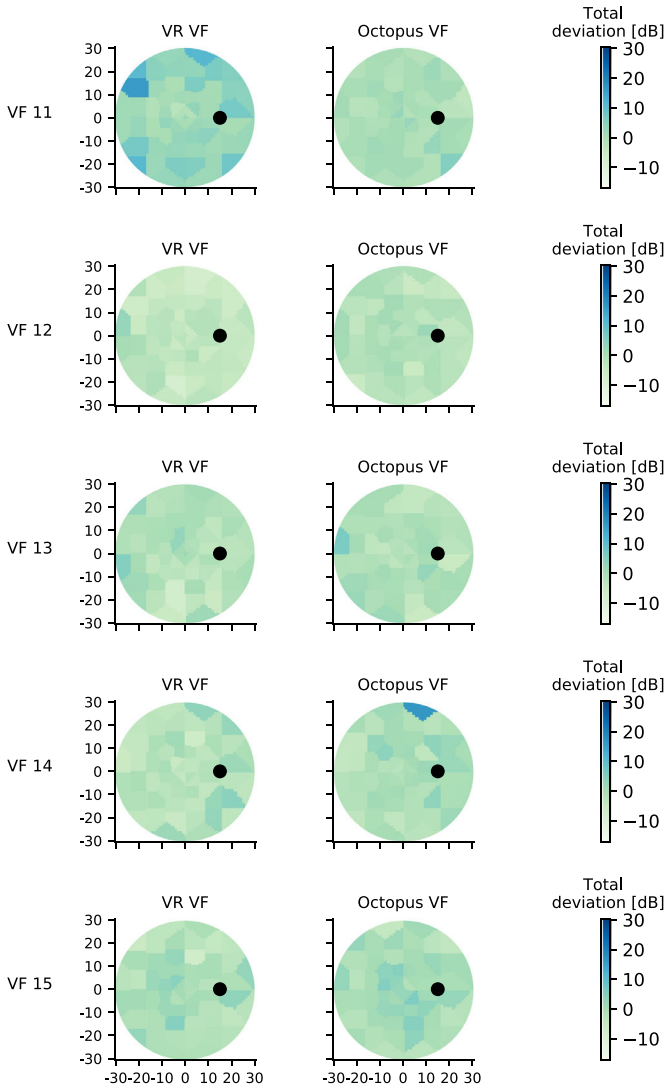

Figure 18. Five healthy VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is, deeper defects.

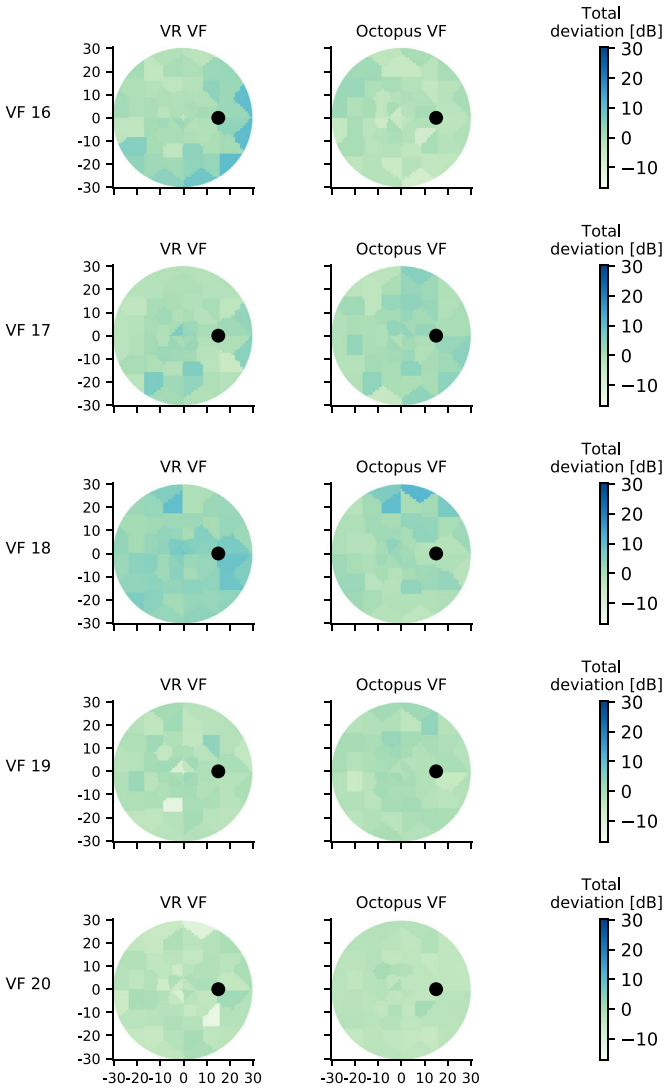

Figure 19. Five healthy VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is , deeper defects.

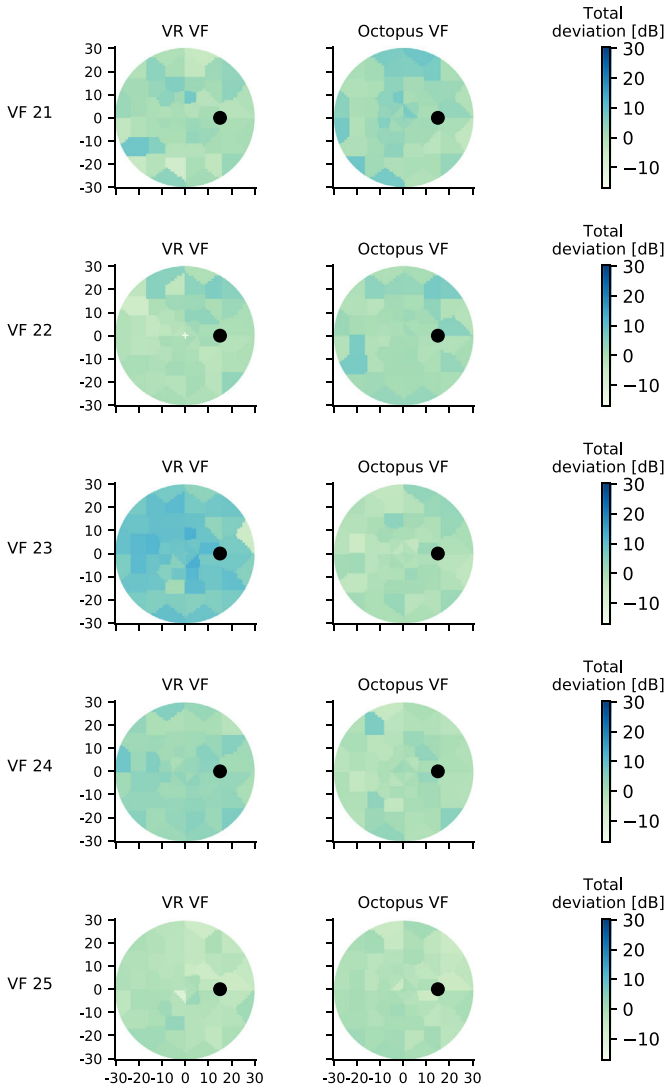

Figure 20. Five healthy VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is , deeper defects.

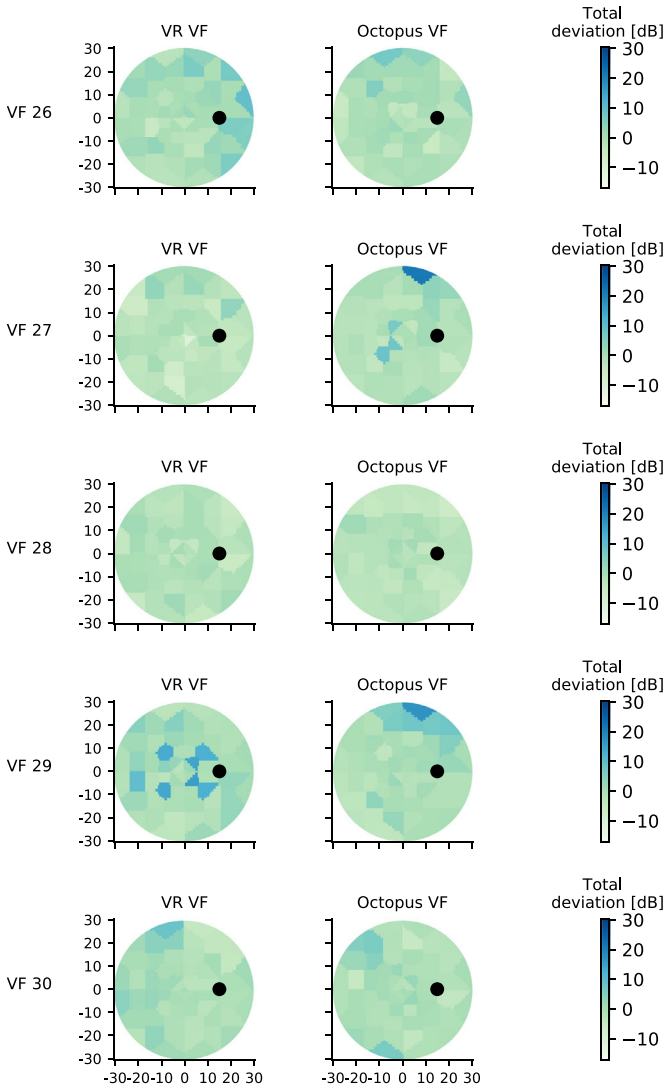

Figure 21. Five healthy VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is , deeper defects.

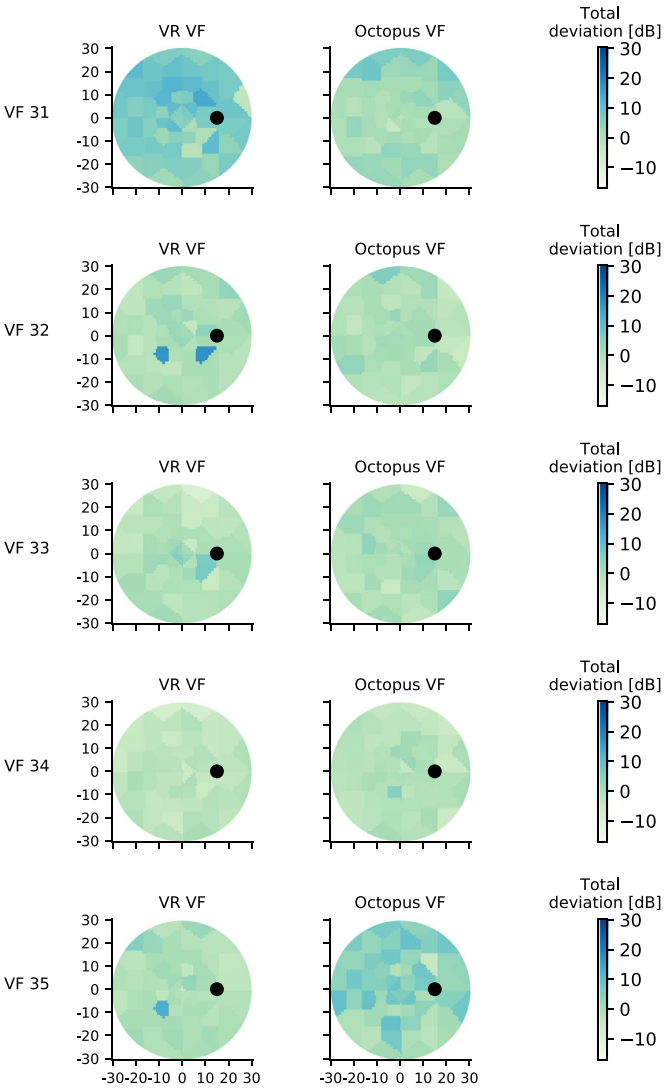

Figure 22. Five healthy VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is , deeper defects.

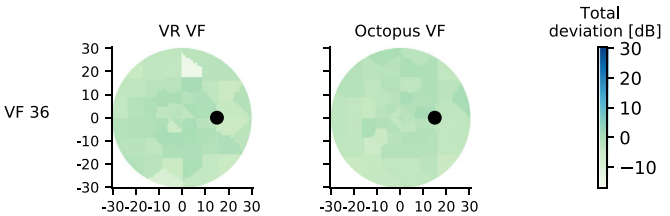

Figure 23. One healthy VF example. It compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is , deeper defects.

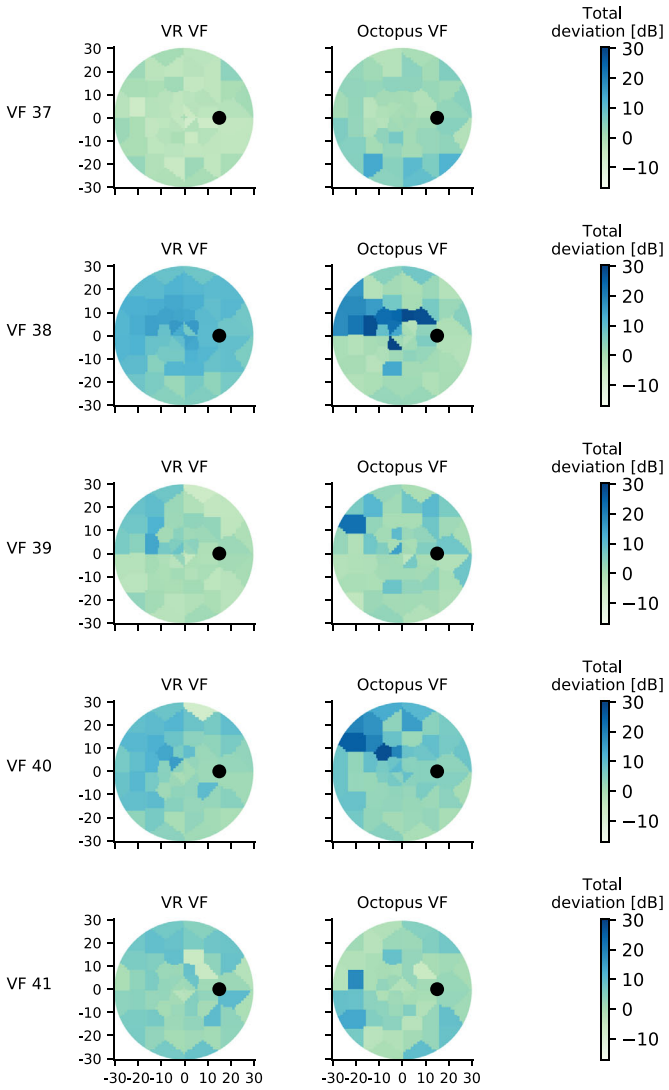

Figure 24. Five glaucoma VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is , deeper defects.

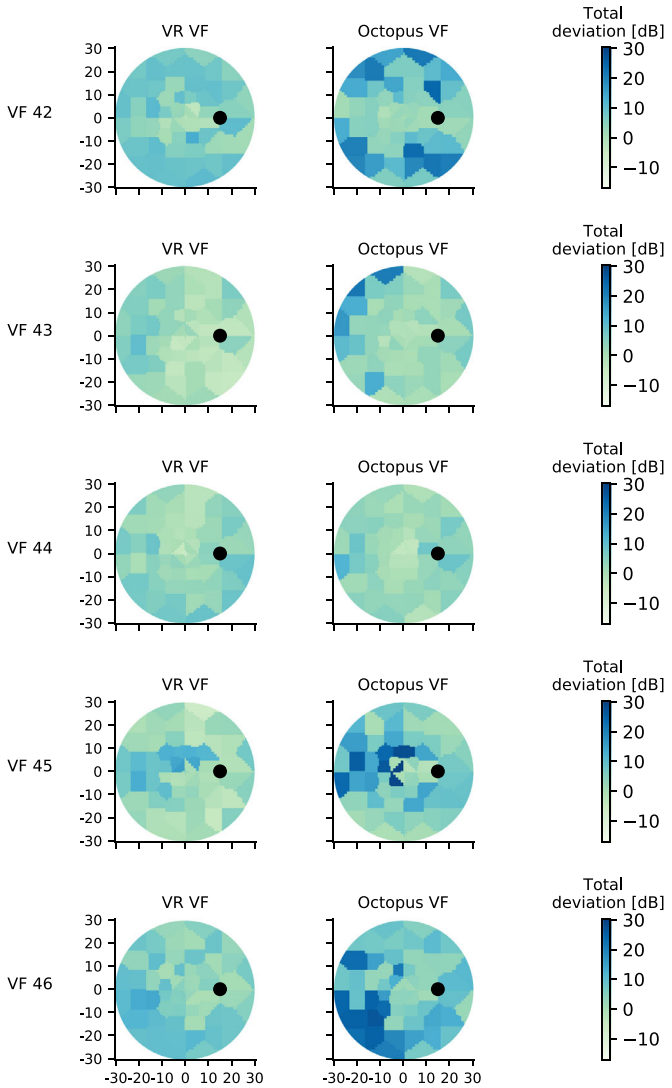

Figure 25. Five glaucoma VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is , deeper defects.

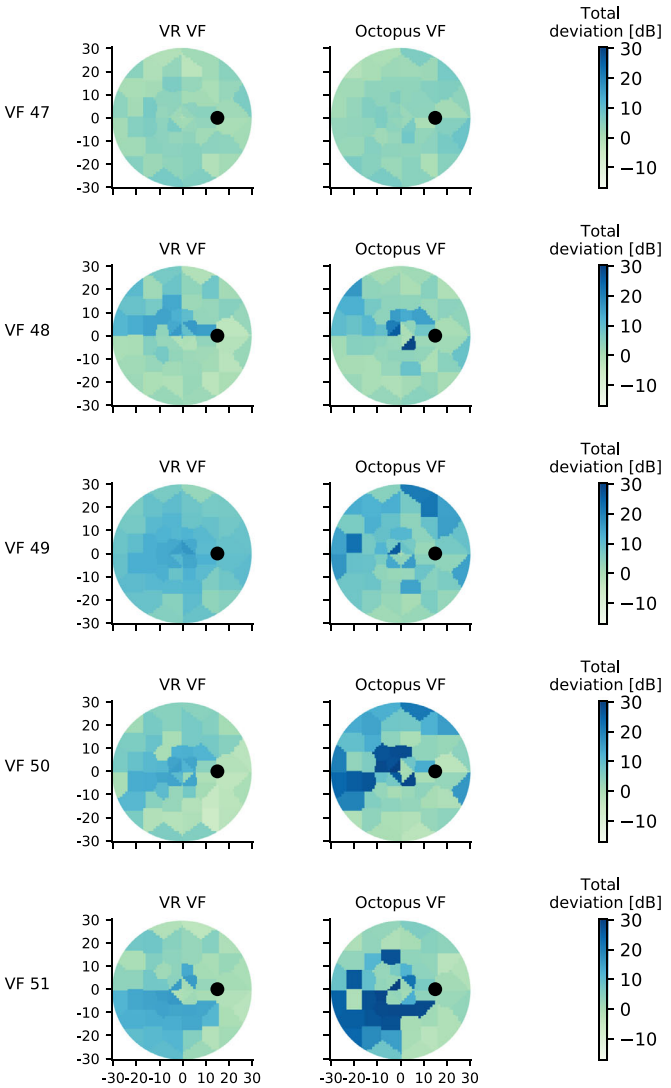

Figure 26. Five glaucoma VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is , deeper defects.

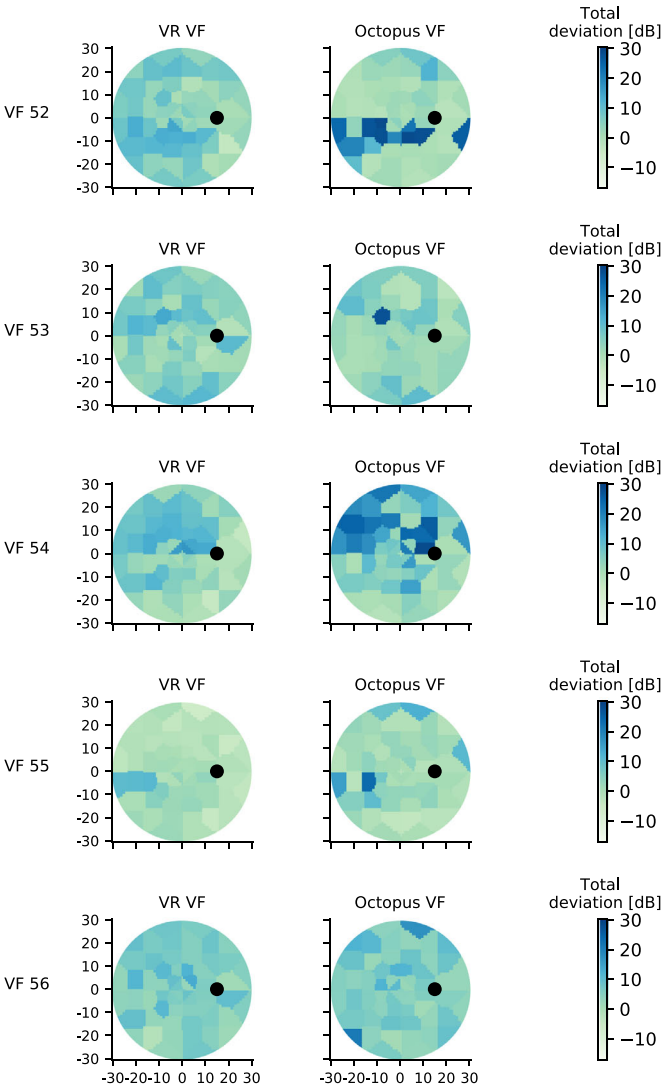

Figure 27. Five glaucoma VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is , deeper defects.

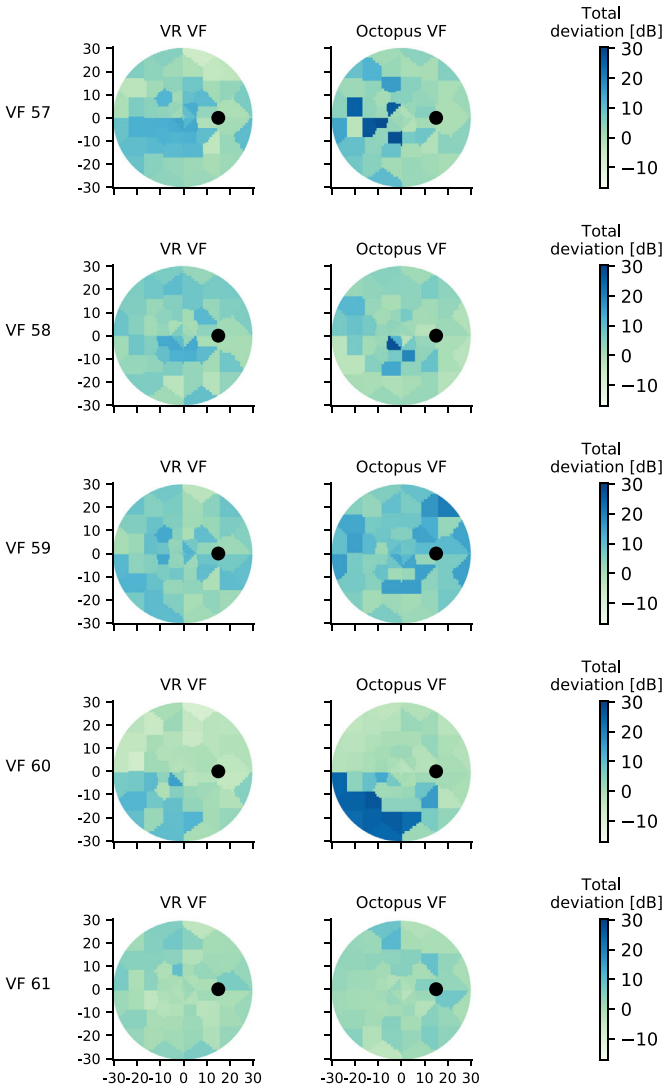

Figure 28. Five glaucoma VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is , deeper defects.

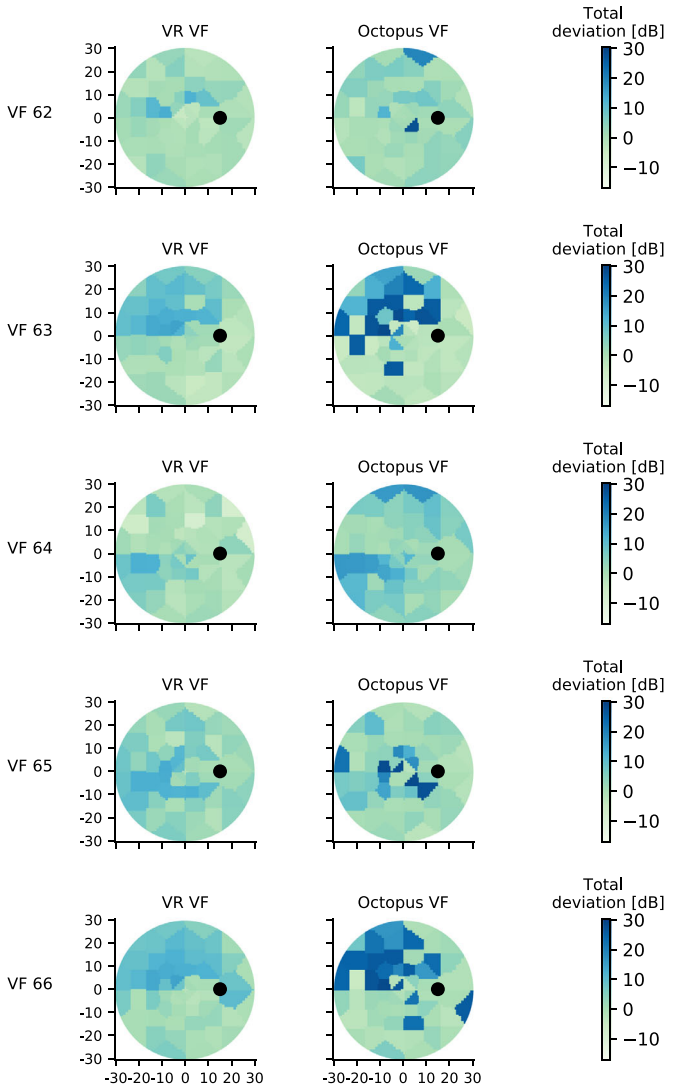

Figure 29. Five glaucoma VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is , deeper defects.

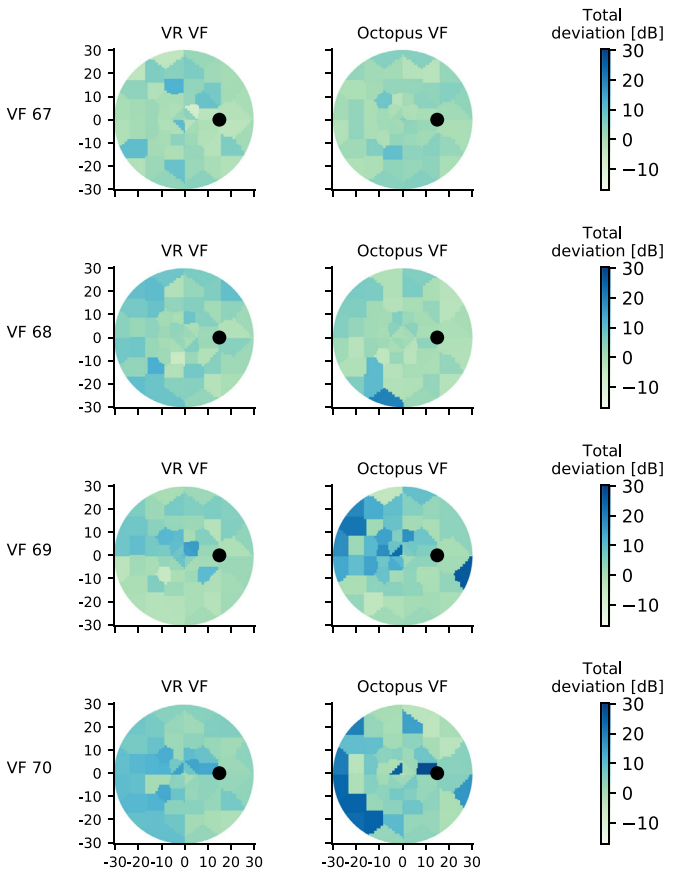

Figure 30. Four glaucoma VF examples. Each row compares acquisitions by VR perimetry (*left*) and by Octopus 900 (*right*). *Black circle* corresponds to the blind spot. *Bluish colors* reflect higher deviations, that is , deeper defects.