Using natural viewing behavior to screen for and reconstruct visual field defects

Birte Gestefeld Laboratory for Experimental Ophthalmology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

Alessandro Grillini Laboratory for Experimental Ophthalmology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

Jan-Bernard C. Marsman Cognitive Neuroscience Center, Department of Biomedical Sciences of Cells and Systems, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

Frans W. Cornelissen Laboratory for Experimental Ophthalmology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

There is a need for simple and effective ways to screen for visual field defects (VFD). Watching a movie is a simple task most humans are familiar with. Therefore we assessed whether it is possible to detect and reconstruct visual field defects based on free viewing eye movements, recorded while watching movie clips. Participants watched 90 movie clips of one minute, with and without simulated visual field defects (sVFD), while their eye movements were tracked. We simulated homonymous hemianopia (HH) (left and right sided) and glaucoma (small nasal arc, large nasal arc, and tunnel vision). We generated fixation density maps of the visual field and trained a linear support vector machine to predict the viewing conditions of each trial of each participant based on these maps. To reconstruct the visual field defect, we computed "viewing priority" maps and maps of differences in fixation density of the visual field of each participant. We were able to classify the simulated visual field condition with more than 85% accuracy. In simulated HH, the viewing priority distribution over the visual field indicated the location of the sVFD in the simulated HH condition. In simulated glaucoma the difference in fixation density to the control condition indicated the location of the sVFD. It is feasible to use natural viewing behavior to screen for and reconstruct (simulated) visual field defects. Movie clip viewing in combination with eye tracking may thus provide an alternative to or supplement standard automated perimetry, in particular in patients who cannot perform the latter technique.

Introduction

The most common way to establish the presence of a visual field defect is by means of standard automated perimetry (SAP), which assesses a person's luminance sensitivity across the visual field. With this test, clinicians can detect and localize visual field defects, already at an early stage of a disease (Wild, 1988). SAP is critical in the management and follow-up of ophthalmic disease. However, it has some requirements and instructions that make it less suitable for use in various groups, such as the very young or very old, or in people with a cognitive impairment. For example, SAP requires the tested person to fixate a small light spot over an extended period of time and to provide behavioral feedback. Both requirements can only be fulfilled if the tested person can remain focused on the task, which is often not achieved (Birt, Shin, Samudrala, Hughes, Kim, & Lee, 1997).

The Rotterdam study (Skenduli-Bala, De Voogd, Wolfs, Van Leeuwen, Ikram, Jonas, Bakker, Hofman & De Jong, 2005) found glaucoma to be the leading cause of visual field loss. Another common cause is stroke,

Citation: 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):11, 1–16, https://doi.org/10.1167/jov.20.9.11.

https://doi.org/10.1167/jov.20.9.11 Received April 24, 2020; published September 17, 2020 ISSN 1534-7362

which can lead to hemianopia or quadrantanopia, where people are blind in one half or one quadrant of their visual field. It also found that the occurrence of visual field defects increases fivefold above the age of 55 years. With an increasing elderly population over the next decades, neurodegenerative eye diseases, such as glaucoma, as well as incidences of stroke, will become more and more common. Because both lead to visual field loss, there is a growing need for fast and easy to perform visual field tests.

A very simple and inexpensive method to assess the visual field is confrontational perimetry. In this method, the examiner presents both index fingers in opposing hemifields, while the patient is watching with one eye and fixating the doctor's opposite eye. The patient has to report which of the index fingers of the examiner wiggled. The sensitivity of this test, however, to detect certain VFDs is very low (Johnson & Baloh, 1991).

If we could find a method that is easier to perform than SAP and more accurate and reproducible than confrontational perimetry, it could open up regular screening to a larger group of patients. With the advances in eye tracking technology, a possible way to screen for VFDs would be to use eye movements. Eye movements are a spontaneous visual behavior and there is ample evidence that the presence of a VFD evokes eye movement behavior that is different from normal (Asfaw, Jones, Smith & Crabb, 2018; Kanjee, Yücel, Steinbach, González & Gupta, 2012; Lamirel, Milea, Cochereau, Duong & Lorenceau, 2014).

For patients with homonymous hemianopia (HH) or glaucoma, it has been shown that the presence of a visual field defect influences eye movement behavior during different tasks. Patients with HH are usually aware of their VFD, and many of them use compensatory strategies in situations like driving or visual exploration during shopping, by making more saccades toward the blind hemifield (Bahnemann, Hamel, De Beukelaer, Ohl, Kehrer, Audebert, Kraft & Brandt, 2015; de Haan, Melis-Dankers, Brouwer & Tucha, 2015; Kasneci, Sippel, Heister, Aehling, Rosenstiel, Schiefer & Papageorgiou, 2014). Differences in eye movement behavior between HH patients and healthy controls can also be observed during reading: patients with HH make more saccades and regressions per line (Trauzettel-Klosinski & Brendler, 1998).

Also, glaucoma patients show viewing behavior different from normal-sighted controls. For example, when glaucoma patients search for target objects in photographs, they make fewer saccades than healthy controls (Smith, Glen, & Crabb, 2012). As glaucoma patients can have defects in many different areas of the visual field and many of them also experience filling-in of their VFD and are not as aware of its location (Crabb, Smith, Glen, Burton & Garway-Heath, 2013), it is harder to define an "appropriate" compensatory eye movement strategy for glaucoma. However, Kübler, Kasneci, Rosenstiel, Heister, Aehling, Nagel, and Papageorgiou (2015) found that some glaucoma patients were able to pass a simulated driving test, and these patients showed increased visual exploration compared to the glaucoma patients who failed the test.

The above-mentioned studies showed significant differences between groups, but a useful screening method needs to be able to distinguish between individual participants. That this is feasible was demonstrated by Crabb, Smith, and Zhu (2014), who used machine learning to classify observers as glaucoma patients or healthy controls and reached a sensitivity of 76% and a specificity of 90%. They used "saccade maps," which were computed from eye movement data collected during free viewing of three different movie clips as an input to a naive Bayesian classifier.

While using machine-learning algorithms can be very successful in distinguishing between different viewing conditions, clinicians usually like to know which part of the visual field is still intact and which is not. The Humphrey field analyzer therefore computes a map of the visual field. An example of this can be seen in Figure 1.

We chose to simulate different full VFDs that occur in HH and in glaucoma, rather than to assess this in actual patients. Simulating the VFDs has the advantage that we know exactly the "ground truth," i.e., the part of the visual field that was masked. This means we can actually verify the quality of the reconstruction of the VFD. However, we simulated only absolute VFDs (no light sensitivity), which in the case of glaucoma usually occur only in later stages of the disease. This means that glaucoma patients usually get used to their loss of vision over time. More gradual visual field loss was not simulated in this study. We will therefore also not aim to compute a map that shows the severity of the visual field loss, as shown in Figure 1, but aim to approximately reconstruct the (simulated) blind area in the visual field.

We know from previous studies, that artificial scotomas have a strong impact on the eye movement behavior of normal-sighted participants, as was shown by (David, Lebranchu, Perreira Da Silva & Le Callet, 2019). In their study participants freely explore static images of different scenes. They are able to identify different eye movement parameters on the basis of which they are able to predict the viewing condition (simulated glaucoma, simulated macular degeneration or control) of the participant in a certain trial. Further evidence for this is that when performing a visual search task with artificial central and peripheral stimuli, normal-sighted participants show increased fixation durations in comparison to performing the task without the artificial scotoma (Cornelissen, Bruin, & Kooijman, 2005). Tant, Cornelissen, Kooijman & Brouwer (2002) find that simulated HH elicits a similar scanning behavior as real HH in participants who perform a dot

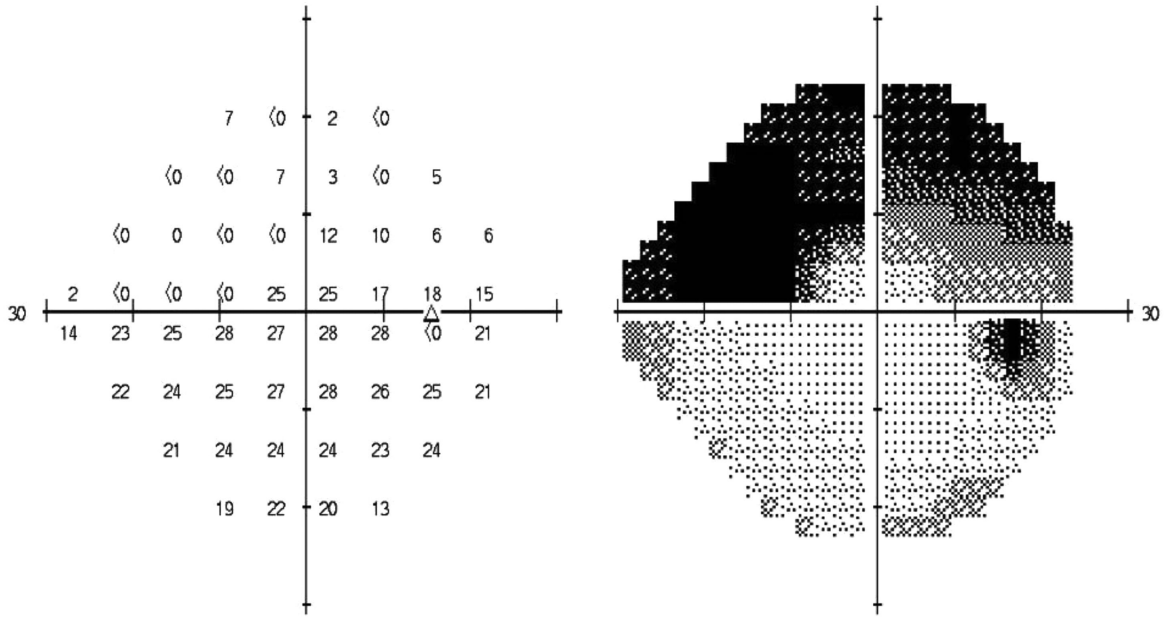

Figure 1. Example of a visual field map computed by the Humphrey field analyzer (HFA). The darker colors indicate a lower contrast sensitivity.

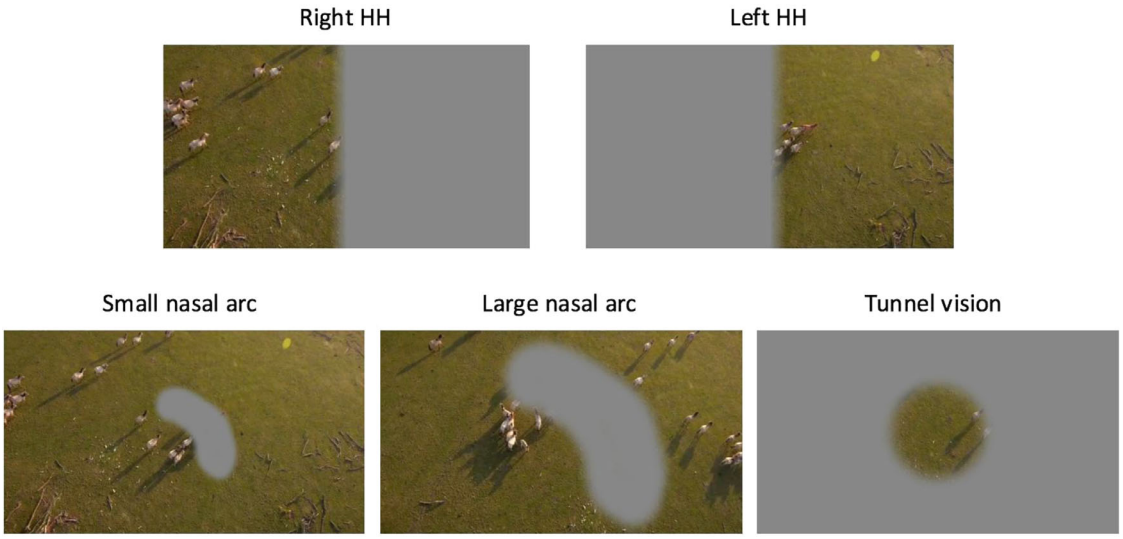

Figure 2. Illustration of the simulated VFDs overlaid over different scenes of the same movie.

counting task. An obvious disadvantage of simulations is that they may not fully comprise the actual experience of people with an actual VFD (Crabb et al., 2013) and that the effects we measure are acute. Consequently, participants have relatively little time to adapt their behavior in response to the VFD.

In our study, participants watched movie clips of one minute's duration, some clips with either simulated HH or simulated glaucoma and the rest of the clips in the control condition. We asked two questions: (1) Can we predict the viewing condition of a specific trial based on fovea-referenced differential fixation density maps? (2) Can we reconstruct the location of the sVFD for each participant? Both questions should be answered affirmatively for the method to be potentially useful in clinical practice.

We used machine learning to determine the accuracy with which we can predict the viewing condition. On

top of that, we obtained maps of the visual field (in analogy to the one provided by SAP) that locate the sVFD for each participant per scotoma archetype.

In addition to answering these two main research questions, we wanted to find out whether the content that was shown had an effect on the eye movement behavior, and more specifically, the possibility to reconstruct the sVFD. So, we obtained visual field maps for each movie clip per scotoma archetype, averaged across observers.

We also aimed to understand the underlying eye movement strategy of the participants in the different sVFD conditions. We computed the average saccade amplitude in 18 different directions for each viewing condition to determine whether a certain scotoma archetype led to a characteristic pattern of saccade amplitudes, which could be an indicator for a compensatory strategy.

Methods

Participants/Study Population

We tested 70 healthy participants (60 females, mean age: 20 years, standard deviation: 3.3 years) with normal or corrected to normal visual acuity and no visual field defect. The ethics committee of the Department of Psychology of the University of Groningen (RUG) approved the study protocol. Participants were psychology students who received study credits for their participation. All participants provided written informed consent. The study followed the tenets of the Declaration of Helsinki.

Materials

The data for this study was collected in two separate experiments, where the first experiment contained simulated HH trials and the second experiment contained simulated glaucoma trials. We analyzed the two resulting data sets separately using the same methods. The sets were collected on two different displays, because set 2 was collected at a later time point, when the equipment at the laboratory had been changed. In set 1, all sVFD were HH (left and right sided). In set 2 all sVFD were glaucoma (small nasal arc, large nasal arc and tunnel vision). The first set was collected with 40 participants, who watched the clips with the simulated HH conditions. We presented movie clips on a 40 cm by 30 cm (1152 × 870 pixel) screen (LaCie CRT). This means that this screen covered a visual field of 36.9° × 28.1° (of visual angle). For set 2, 30 different participants watched movie clips on a 50 cm × 35 cm (1920 × 1080 pixel) screen (BenQ Zowie xl2540). The second screen thus covered a visual field of 45.2° × 32.5°. For both sets, monocular eye movements were recorded with an Eyelink 1000 eye tracker (SR Research) at 1000 Hz, which was connected to a dedicated PC with an ethernet cable running Eyelink software. At the beginning of the experiment, we tested whether one eye calibrated more accurately or could be tracked more robustly. If this was the case, we chose to track this "better" eye. Otherwise we randomly chose the eye to be tracked. The host PC was connected to a laptop running MATLAB (Version 2016b, MathWorks, Natick, MA) with the Psychtoolbox (Brainard, 1997; Kleiner, Brainard, Pelli, Broussard, Wolf & Niehorster, 2007) and the Eyelinktoolbox (Cornelissen, Peters, & Palmer, 2002) via Ethernet. All participants were seated at 60 cm distance from the screen.

Stimuli

All movie clips were presented on the full screen. The sequences for both experiments were taken from 18 different movies, which could be grouped into the following categories: animations, cartoons, feature films (crime stories, dramas, and comedies), and nature documentaries. The movie clips contained a wide range of different content. While some showed actors in different contexts others showed animated characters or landscapes and animals. For a complete list of movies from which we extracted the clips see the supplementary material. The original language of all movies was Dutch. Stimuli were presented using MATLAB with Psychtoolbox and gstreamer (https://gstreamer.freedesktop.org). Observers watched the clips without any task instructions, while their head was placed on a chin rest. The visual field defect was simulated by masking a part of the movie with gray bitmaps, which are shown in Figure 1. For set 1 each participant watched 90 movie sequences in total. Thirty or 60 movies were presented with simulated HH, either on the left or on the right side. For set 2, each participant watched 88 movie clips, 66 of these clips with a sVFD: a small nasal arc, a large nasal arc and tunnel vision (22 clips each sVFD). All movie clips were presented for 1 minute.

Auditory content was presented over headphones (Philips Ear clip headphones SHS4700/10). Volume was adjusted so that it was audible and comfortable for the participant.

Simulations

The sVFD were created as grayscale bitmaps in MATLAB (simulated HH) and GIMP (simulated glaucoma), which were overlaid over the current video frame. The simulations were coupled to the eye tracker

and the mask moved along with the participant's gaze, always covering the same part of the visual field. We presented the video frame as well as the mask as Psychtoolbox textures, which had the same refresh rate, namely 84 Hz for the simulated HH conditions and 60 Hz for the simulated glaucoma conditions. So, before each frame was drawn the program checked the gaze position and updated the location of the mask, accordingly. As the sampling rate of the eye tracker was higher than the refresh rate of the stimulus, the position of the mask could be updated in real time. If there was no sample available (e.g. because of a blink) the last available gaze position was used to position the mask. When participants gazed towards the center of the screen, the mask in the simulated HH conditions covered 50% of the screen horizontally and the entire screen vertically. In the glaucoma conditions, the small nasal arc covered 4% of the screen having a width of approximately 5° and a length of about 15°. The large nasal arc covered about 15% of the screen having a width of 12° and a length of about 27°. The tunnel vision simulation covered 86% of the screen masking everything outside the radius of the inner 15°.

Due to the gaze contingency of the simulations, depending on the momentary gaze direction of the participant, the masks could cover a larger or smaller part of the screen throughout a trial. In the simulated HH conditions, for example, up to 100% of the screen could be masked, if participants looked toward the nonmasked side of the screen.

Procedure

Participants were seated in front of a computer screen with their head fixed in a chin rest. We calibrated the eye tracker once before starting the experiment using the built-in calibration procedure of the Eyelink. Participants watched either 90 or 88 movie clips of one minute's duration. As an initial goal of the first experiment was to test the influence of presenting audio content from different directions on the viewing behavior, the participants had to change the ear phones after each trial. The auditory condition was indicated on the screen before the experimenter started the trial. Sound was presented to the left or right or both ears or was muted. In addition, the screen showed information for the experimenter which keys should be pressed to continue or repeat the trial, or to exit the experiment. In both experiments, the next trial started if the experimenter pressed the "spacebar." Participants were informed that they could take a break in between trials if they felt tired. The experimenter also asked after one third and two thirds of the trials if the participant wanted to take a break. After breaks, we performed a drift correction. Moreover, if the participant changed their seating position or if we noticed that the eye tracker lost the gaze position frequently, we recalibrated the eye tracker.

The order in which the movie clips were presented was randomized for each participant.

Eye-movement analysis

The Eyelink 1000 gives an average eye position accuracy of better than 0.5°. The average calibration error was <1° and the maximum calibration error <1.5°. Participants who could not be calibrated accurately were not included. Data of three participants from the first set, and data of two participants from the second set, were excluded due to an inaccurate calibration (average calibration error above 1° and maximum calibration error above 1.5°).

Saccades and fixations were defined using the Eyelink's built-in algorithm. Fixations separated by a blink were concatenated.

Classification

Our first aim was to discriminate between viewing conditions based on the eye movement data. To do so, first, we extracted all fixations from the data. Next, we defined the center of the visual field to be the position of the first fixation. The position of the second fixation is defined with respect to that of the first fixation, and so on. In this way, we follow the visual field of the participant, always defining the position of a fixation with respect to the previous one.

We then defined grids over the visual field to compute the proportion of fixations that fell into each bin of the grid in one trial. As the two screens had different sizes, we based the size of the grid on the average saccade amplitudes of all participants in the control conditions of each screen. We computed the average saccade amplitude and added two standard deviations in 18 different directions (see Figure 3a). We then defined a rectangular grid that covered all fixations that fell into this area. For set 1 the grid covered an area of 30° × 28° and for set 2 the grid covered 44° × 28°. In both sets the bin size was 2° in horizontal and vertical direction, which is the same as the resolution of the grid used by Crabb et al. (2014). For each trial, we computed the sum of fixations per bin and z-normalized the resulting map (see Figure 3b for examples).

The maps were vectorized and used to train a linear support vector machine to distinguish between the control and sVFD conditions, which were in set 1 control, right-sided HH, and left-sided HH, and in set 2 control, small nasal arc, large nasal arc, and tunnel vision. The classifier was trained and tested by means of a 10-fold cross validation, with two thirds of the data used as training data. We firstly distinguished

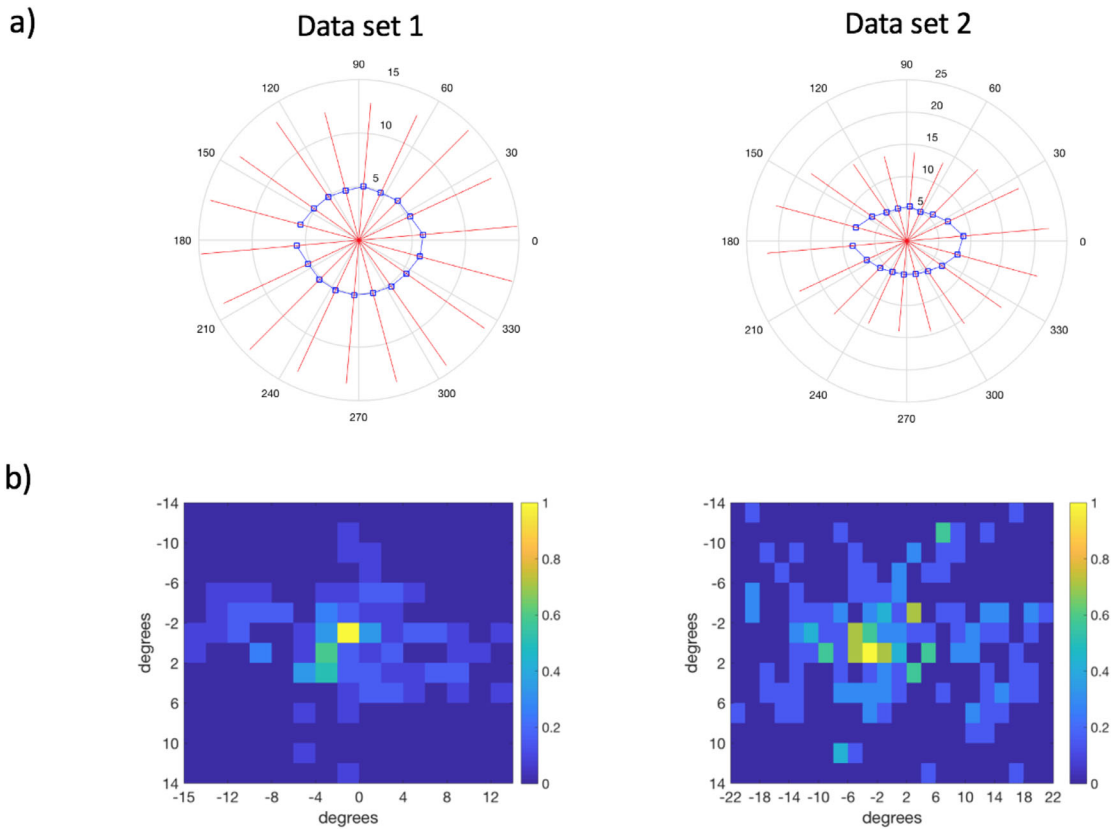

Figure 3. (a) Saccade amplitudes averaged over all participants in the control condition. The blue squares show the mean saccade amplitudes per direction and the red lines are the corresponding error bars showing 2 standard deviations. (b) Examples of fixation maps of one trial from each data set showing the proportion of fixations that fell into each bin.

between the control condition and the presence of an sVFD (without making a distinction between different scotoma archetypes) to determine the sensitivity and specificity, which we visualized as an ROC curve. Second we determined the classification accuracy per sVFD archetype and visualized the results as confusion plots.

Because data set 1 contained different audio conditions, we wanted to exclude the possibility that there was an effect of the different auditory conditions on the viewing behavior of our participants. We therefore performed the same classification procedure as described above, but selecting trials from the (visual) control condition and distinguishing between the four different audio conditions (stereo, left ear, right ear, and no sound). Because the classification accuracy was chance level, we concluded that different auditory conditions did not influence viewing behavior and did not distinguish between auditory conditions for our further analysis.

Reconstruction of the VFD

For the reconstruction of the visual field, we collected all trials that were performed with the same sVFD by each participant. We then based the reconstruction of the respective sVFD on this set of trials.

We used two different approaches to reconstruct the simulated VFD on both data sets. In our first approach, we used the concept of viewing priority (Marsman, Cornelissen, Dorr, Vig, Barth & Renken, 2016).

Viewing priority (VP) measures how consistent the viewing behavior of one observer is to that of other observers based on fuzzy c-means clustering. The value of VP is measured as the distance of a given fixation to a set of reference fixations, which are fixations made by other observers watching the same context (i.e., a movie scene in this experiment). These reference fixations usually cluster around the same aspect of the scene. To determine how closely clustered the reference fixations are, a set of random fixations is selected, which are fixations made by other observers watching different content (i.e., a different movie scene), and the distance of the reference fixations to the random fixations is computed. The height of the VP measure is dependent on the combination of the density of the reference fixation and the distance of the observers' fixation to these clusters. For a detailed explanation of the calculation of VP, we refer to Marsman et al. (2016).

The VP for each fixation was computed by selecting reference and random sets from fixations made under normal viewing conditions for the same screen size. The size of the reference set and random set differed between fixations, as other observers may not make a fixation during a certain time interval. The set of reference fixations had a size of at least 10 fixations and the random set was three times the size of the reference set. We show how the VP is distributed across the visual field by first computing a heat map, where each fixation is represented as a Gaussian distribution with a width of 1°, weighted by its VP value. We then divide this map by the map that shows the distribution of fixations (without weights).

In our second approach, we used differences in fixation density between the control condition and the simulated VFDs followed by permutation statistics to test for significant differences. The following analysis steps were used to reconstruct the visual field defect:

(1) Per participant, we calculated a reference map for the mean fixation density in the control condition by averaging across all other participants' control condition trials.
(2) Computation of a mean fixation density map for each viewing condition per participant (VFD map).
(3) We calculated a difference map d by subtracting the reference map from each VFD map.
(4) For each participant, compute the null distribution of differences by a) permuting n times across both control trials and simulated VFD trials and b) generating a set of random difference maps R, by drawing two random samples from the permuted sets of trials and subtracting the two.
(5) The probability map is then calculated as follows: $P(x,y) = \text{sum}(d > R)/n$.

For each observer, we applied both approaches and computed the VP map and the differential fixation map for each condition. In addition, we also created average VP and differential fixation maps across all observers for each condition.

To calculate the performance for each approach, we calculated the correlations between both the VP map and the differential fixation map, and the binarized map of the visual field where the area of the defect is masked.

Both approaches were also used to compute maps for each condition and each film averaged over all participants, who had watched the same clip under the same viewing condition.

Eye movement strategies

As the last part of our data analysis, we investigated whether participants developed particular eye movement strategies when watching the movie clips with a certain sVFD archetype and whether they changed their behavior over time. We computed the average saccade amplitude in each viewing condition for each participant in 16 different directions, together spanning the full 360°. In addition, to examine the possible presence of learning effects, we computed the average saccade amplitude and fixation duration of all participants per viewing condition over the first, middle and last tercile of all trials. To assess whether these eye movement features differed among these three sets, we performed a one-way analysis of variance (ANOVA).

Results

We will first describe the results of the classification of each set into control and sVFD conditions, as well as into the sVFD archetypes using the distribution of fixations in the visual field as an input. Next, we show how well the VP and the relative fixation density maps reconstructed the sVFD of each sVFD archetype. Finally, we will present the saccade amplitude per direction for each type of sVFD averaged across participants.

Classification based on the distribution of fixations

Assuming that this method will be used to screen for an underlying pathology, we first aimed to distinguish between the control condition and the presence of an sVFD. In addition, we wanted to find out whether it is also possible to distinguish between different sVFD archetypes within each data set to test the limitations of the method.

Figure 4b shows the ROC curve for both parts. The area under the curve for set 1 was 0.91. The area under the curve for set 2 was 0.87. Figure 4c shows the confusion plot with the mean percentages of trials labeled as each sVFD archetype. The overall average accuracy for archetype classification in set 1 was 85.55% (minimum: 83.93%, maximum: 86.86%). The overall average classification accuracy for archetype classification in set 2 was 85.81% (minimum: 84.11%, maximum: 88.14 %). The confusion plots show that the classification accuracy for the control condition is the highest with over 92% correctly classified trials. The lowest classification accuracy was achieved for the small nasal arc condition, for which 23.63% of trials get misclassified as being the control condition.

Reconstructing the sVFD

In the following two sections, we will show the results of the reconstruction of the sVFD using the

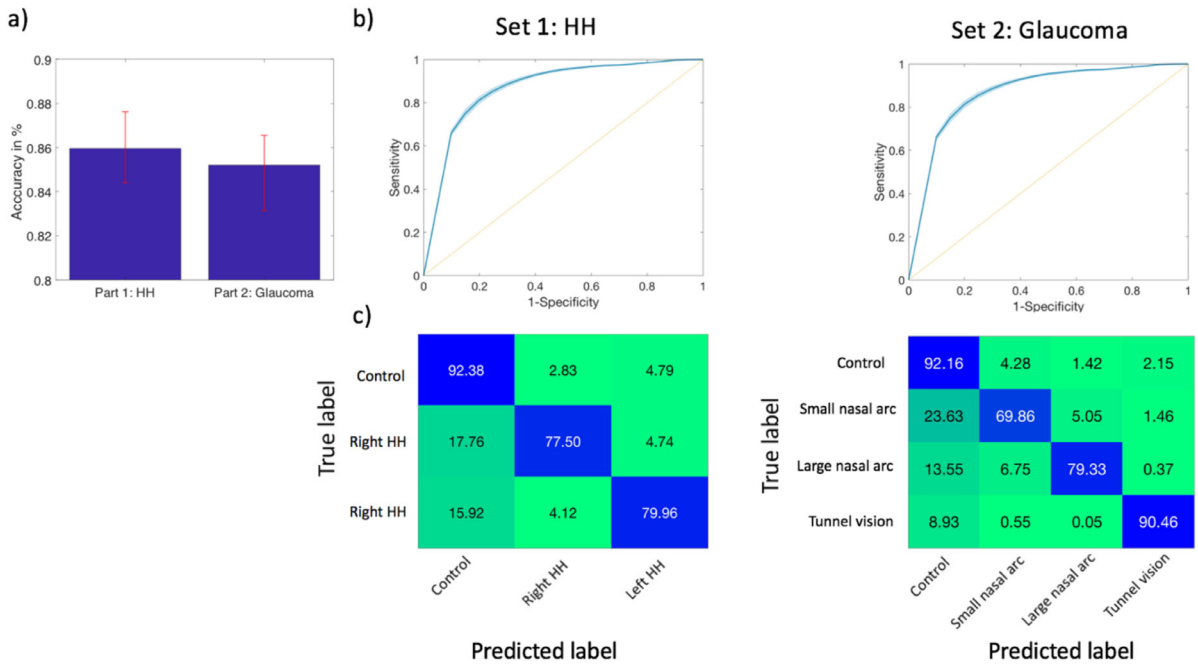

Figure 4. (a) Overall classification accuracy of each part with error bars showing the minimum and maximum accuracy of the 10-fold cross-validation. (b) ROC curves for detecting the presence of an sVFD. The area under the curve for set 1 is 0.91 and for set 2 is 0.87. Error bars show the 95% confidence interval. (c) The confusion plots when distinguishing between the different sVFD archetypes.

distribution of VP and the comparisons of fixation density in the visual field per participant.

We found that using the VP maps is our best approach to reconstruct the simulated HH archetypes, whereas the differential fixation maps are more closely correlated with the simulated glaucoma archetypes.

In the third section, we will show the results of the reconstruction based on both approaches for specific movie clips. The correlation of the reconstruction with the actual sVFD varies between movie clips.

Reconstruction of the sVFD based on VP

Under normal viewing conditions, VP is high in all regions of the visual field, while in the simulated HH condition, the areas that are visible correspond to areas where fixations have a high VP (close to 1) and the masked areas correspond to areas with fixations that have on average a low VP (close to 0). The average correlation coefficient between the simulated right sided HH and its reconstruction was 0.38 (standard deviation [SD] = 0.23). The average correlation coefficient between the simulated left sided HH and its reconstruction was 0.24 (SD = 0.31). Figure 5 shows the distribution of VP in the visual field averaged across all participants for each HH condition and some examples of the distribution of VP in the visual field of single observers. The different participants show some variance in the distribution of VP across the visual field. In the simulated glaucoma condition, it does not become clear where the sVFD was located when plotting the distribution of VP. It is simply high across the entire visual field in all conditions (see Figure 6).

Reconstruction of the VFD based on differential fixation density maps

In the differential fixation density maps, a lower p-value indicates a larger difference in the fixation density between the control condition and the simulated VFD conditions. In the simulated glaucoma conditions the distribution of p values averaged across all observers showed that low p values correspond mainly to the masked areas in the visual field. It is also possible to infer the location of the sVFD comparing fixation density of single participants to the average fixation density of all other participants in the control condition. There is some variance in the accuracy of the reconstruction between individual participants, as can be seen in Figure 7. The average correlation coefficient of the small nasal arc was 0.26 (SD = 0.1). The average correlation coefficient of the large nasal arc was 0.36 (SD = 0.11) and the average correlation coefficient of the tunnel vision was 0.58 (SD = 0.18).

In the simulated HH conditions high and low p-values are randomly distributed across the visual field and are therefore not useful to reconstruct the VFD, as shown in Figure 8.

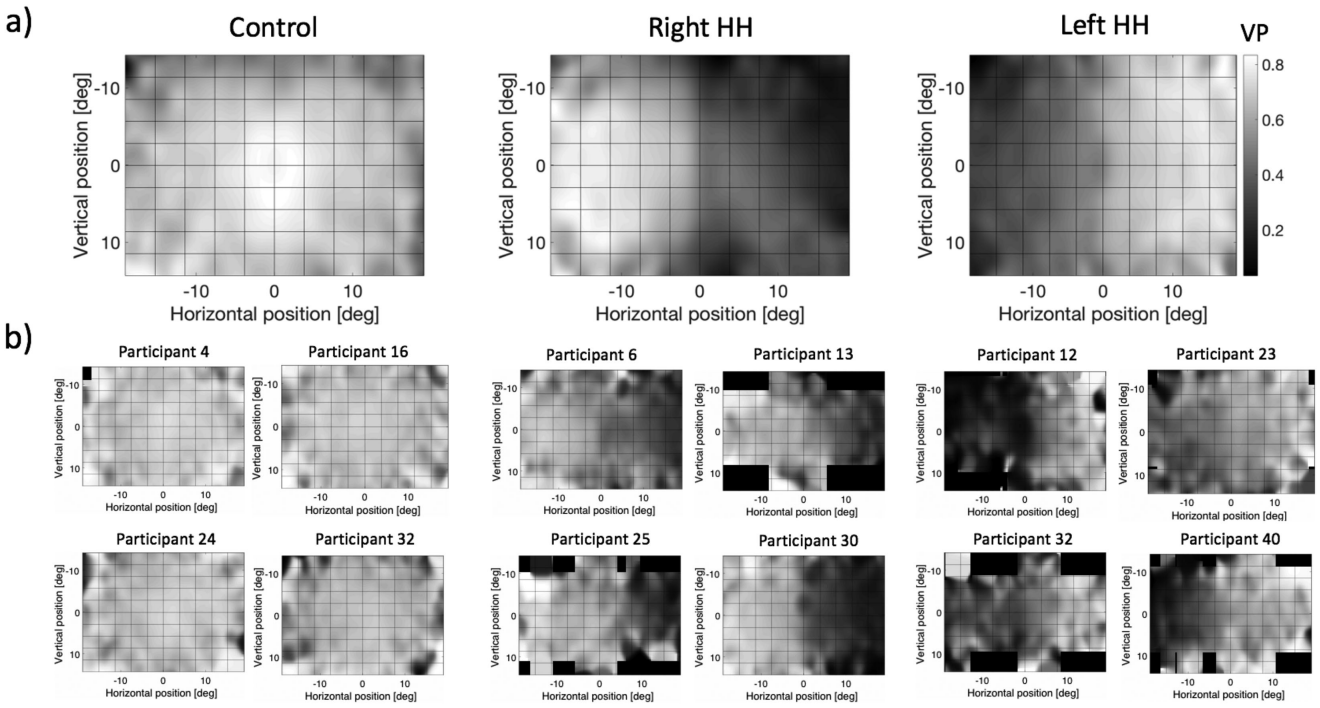

Figure 5. Distribution of VP in the visual field in the simulated HH conditions (a) averaged over all observers (b) examples of individual observers. A lower VP is indicated by a darker shade of gray, showing that the VP is lower in the areas of the sVFD. The grids were added on top of the maps to facilitate orientation for the viewer. The spatial distribution of VP is computed in a continuous manner.

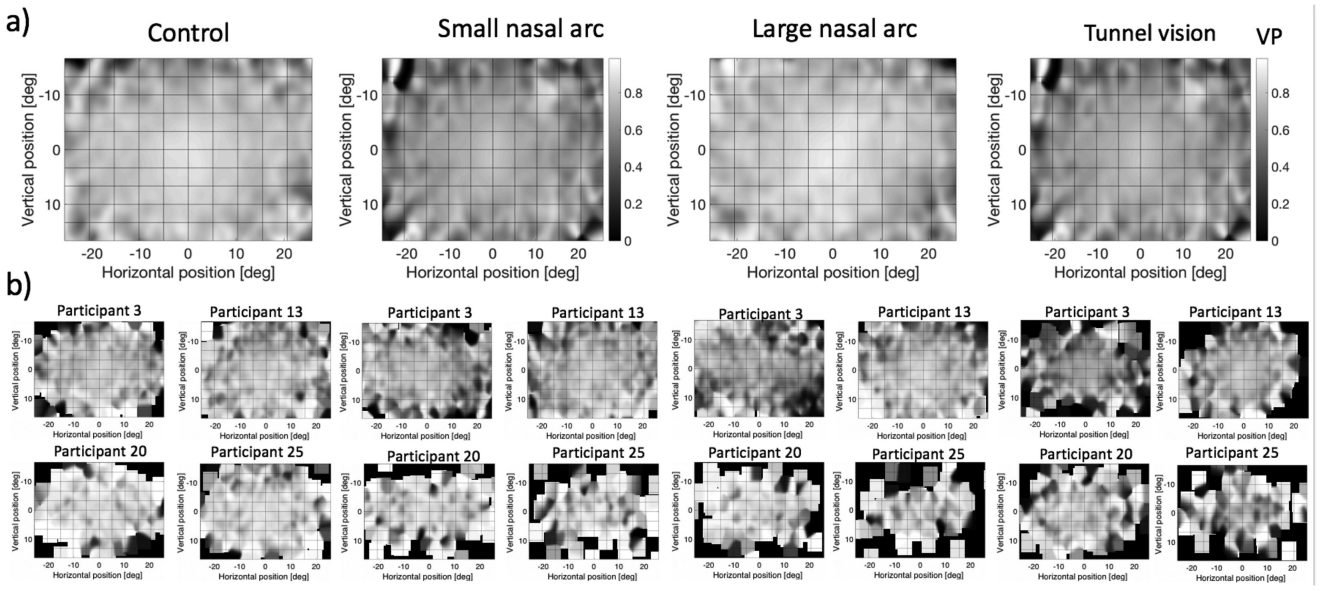

Figure 6. Distribution of VP in the visual field in the simulated glaucoma conditions (a) averaged over all observers (b) examples of individual observers.

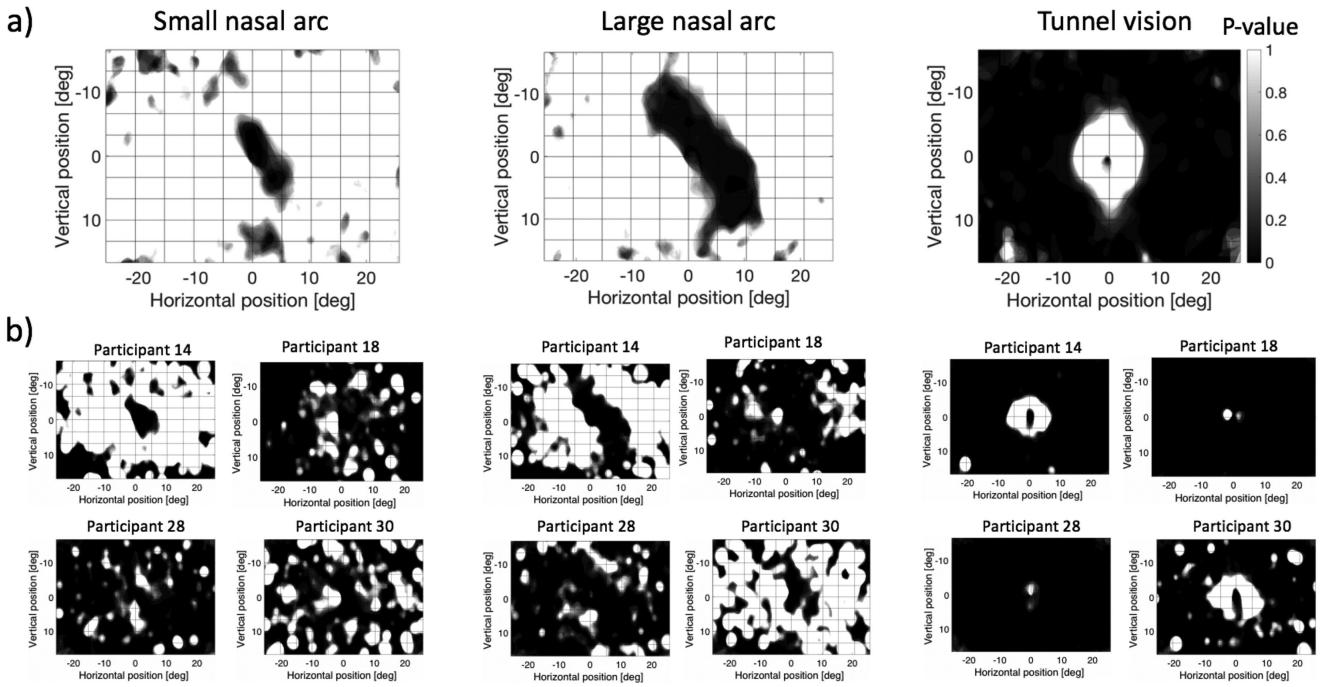

Figure 7. Differential fixation density maps, showing the distribution of *p* values in the visual field in the simulated glaucoma condition (a) averaged over all observers per condition and (b) examples of results obtained with different participants. A darker shade of gray represents a lower *p* value showing that the fixation density in these areas is significantly different from the control condition.

Saccade amplitudes and fixation duration

We computed the distributions of saccade amplitudes in 18 different directions to get a better understanding of how each sVFD influences the eye movements of the participants. In more concrete terms, we were interested in signs for potential compensatory strategies of our participants. Figure 9 shows that in the two simulated HH conditions, participants made on average larger saccades toward the masked hemifield, especially in the horizontal direction.

In the simulated glaucoma conditions, the saccade amplitude depends on the sVFD archetype. In both nasal arc conditions, our participants showed on average higher saccade amplitudes toward the masked parts of the visual field, with a bias for larger saccades in the horizontal direction. In the tunnel vision condition, saccade amplitudes in all directions were decreased compared to the control condition as measured with a two sample *t*-test ($p < 0.001$).

In Figure 10 we show the average saccade amplitudes and fixation durations of the first, middle and last tercile of the trials per viewing condition. We did not find significant differences between the three different groups with a one-way ANOVA, which we calculated for each part.

Tables 1 and 2 show the group means of each viewing condition and the F-values, degrees of freedom and p-values of the ANOVA.

Discussion

Our main conclusion is that it is feasible to predict the presence and archetype of an sVFD with which an observer had been viewing a movie clip, based on the recorded eye movement data, for both simulated field defects that in some ways resemble those seen in HH and glaucoma. For single one-minute trials, this could be done with a mean accuracy of 85.6% (simulated HH) and a mean accuracy of 85.8% (simulated glaucoma). In addition, we were able to reconstruct the sVFD based on data from the same experiments, provided we integrated the data of at least several one-minute recordings. In fact, when combining the data of all participants that performed a specific simulation, visual field reconstructions compared quite well to the simulated defects. Our results imply that the presence of a (simulated) visual field defect sufficiently changes viewing behavior (of normal-sighted controls) to use eye movement data as the basis for a clinically relevant

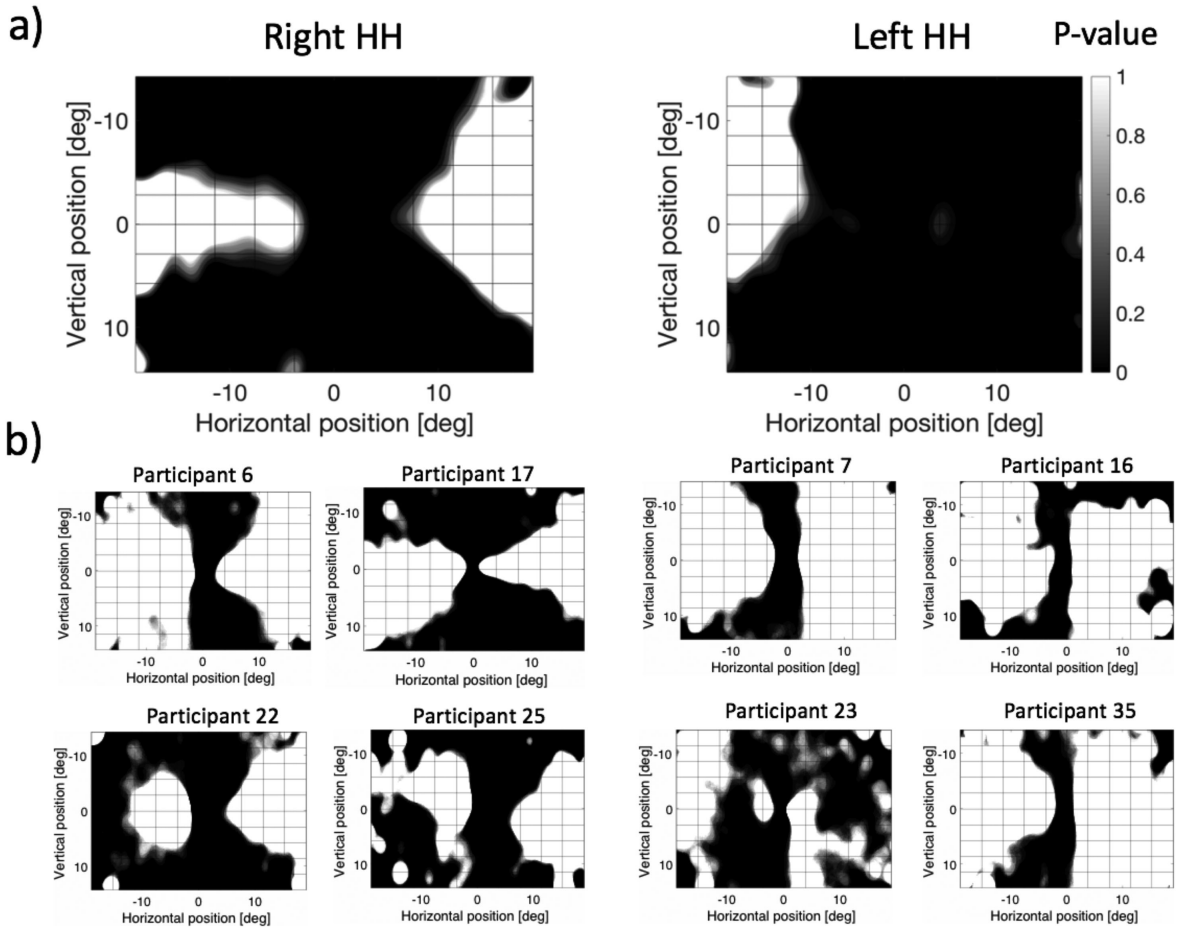

Figure 8. Differential fixation density maps, showing the distribution of *p* values in the visual field in the simulated HH condition (a) averaged over all observers per condition and (b) examples of different observers.

screening on visual field defect, e.g., in patients unable to perform standard automated perimetry. In contexts in which the assessment time is not a limitation (e.g., at home), reconstructions may even become quite accurate.

Machine learning can predict the simulated visual field defect

In line with previous studies (Crabb et al., 2014; David et al., 2019), we found that machine learning is an effective way to separate sVFD from control conditions. We found that the spatial distribution of fixations in the visual field provides the necessary information to separate the sVFD and control conditions.

Based on the confusion plots, we found that the most difficult archetype to detect was the small nasal arc, which is not surprising, as it was the smallest sVFD. The identification of sVFD improved with size, with the tunnel vision condition being the most often correctly identified sVFD.

Especially in glaucoma, it is crucial to detect a VFD early on, when it is small to stop the progression of the disease early on. But it is the most challenging condition to detect.

Because the eye movement behavior is influenced, apart from the type of sVFD, by the content of the movie clip that is shown, we believe that classification accuracy could further improve if we could find the optimal content for the video clips.

Previous studies concluded that saccades are more informative than fixations when attempting to classify eye movement behavior (Crabb et al., 2014; David et al., 2019). In our study, we use the relative distribution of fixations in space ("differential fixation maps"). By necessity, these are, however, closely related to "saccade maps," such as used by Crabb et al. (2014). After all, participants must have made a saccade prior to a fixation. In contrast to saccade maps, the fixation maps also quantify any center bias of an observer, since we do not exclude the center of the visual field from the analysis. Saccade maps, on the other hand, may be more robust with respect to calibration errors as precise landing positions are less relevant.

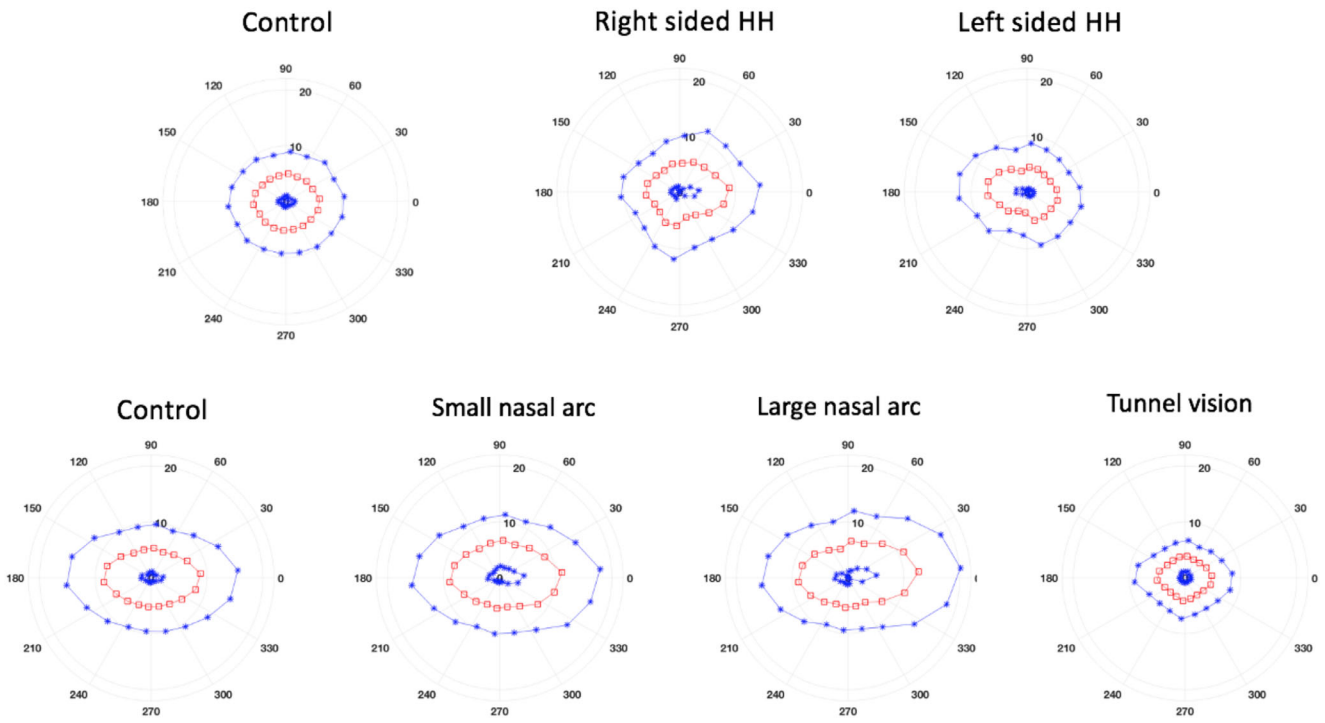

Figure 9. Average saccade length over all participants in the control condition and per sVFD archetype of both parts. The red squares show the mean saccade amplitude and the blue stars one standard deviation of each direction. Saccade amplitudes in degrees of visual angle.

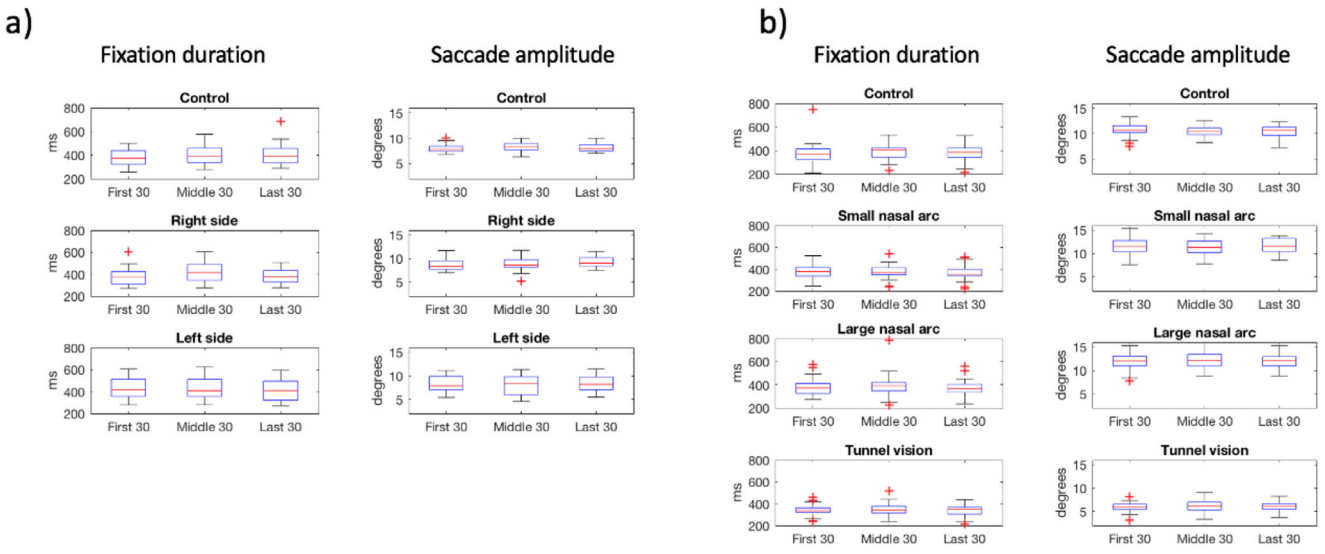

Figure 10. Box and whisker plots of fixation duration and saccade amplitude in each viewing condition a) of set 1 and b) of set 2 for the first, middle and last tercile of all trials. The red line indicates the median value per group, the box indicates the 25th and 75th percentile, and the error bars show the most extreme values. Outliers are shown as red crosses.

Viewing condition	Fixation duration (ms) group means: first, middle, and last tercile of trials	ANOVA results fixation duration	Saccade amplitude (deg.) group means: first, middle, and last tercile of trials	ANOVA results saccade amplitudes
Control	380; 398; 403	$F = 0.56$, $df = 2$, $p = 0.58$	8.0; 8.3; 8.1	$F = 1.05$, $df = 2$, $p = 0.36$
Right sided HH	383;412; 383	$F = 0.51$, $df = 2$, $p = 0.61$	8.9; 8.8; 9.3	$F = 0.48$, $df = 2$, $p = 0.62$
Left sided HH	435; 428; 413	$F = 0.48$, $df = 2$, $p = 0.86$	8.3; 8.1; 8.4	$F = 0.08$, $df = 2$, $p = 0.92$

Table 1. Group means and *F* values, degrees of freedom and *p* values as result of a one-way ANOVA of fixation duration and saccade amplitude of each viewing condition of the first, middle, and last tercile of trials of dataset 1.

Viewing condition	Fixation duration (ms) group means: first, middle, and last tercile of trials	ANOVA results fixation duration	Saccade amplitude (deg) group means: first, middle, and last tercile of trials	ANOVA results saccade amplitudes
Control	354; 365; 357	$F = 0.18$, $df = 2$, $p = 0.83$	10.0; 9.7; 9.7	$F = 0.34$, $df = 2$, $p = 0.71$
Small nasal arc	360; 350; 340	$F = 0.8$, $df = 2$, $p = 0.45$	10.7 deg, 10.7 deg, 10.8	$F = 0.05$, $df = 2$, $p = 0.95$
Large nasal arc	354; 369; 350	$F = 0.8$, $df = 2$, $p = 0.62$	11.2; 11.3; 11.3	$F = 0.08$, $df = 2$, $p = 0.92$
Tunnel vision	318; 323; 315	$F = 0.18$, $df = 2$, $p = 0.83$	5.6; 5.8; 5.7	$F = 0.25$, $df = 2$, $p = 0.78$

Table 2. Group means and *F* values, degrees of freedom, and *p* values as result of a one-way ANOVA of fixation duration and saccade amplitude of each viewing condition of the first, middle, and last tercile of trials of dataset 2.

The location of the sVFD can be reconstructed based on eye-movement behavior

In addition to classifying the viewing condition and sVFD type, we found that it was possible to reconstruct the VFD. Simulated HH could be reconstructed based on the distribution of VP, while the simulated glaucomatous VFD could be reconstructed based on differences in fixation density in the visual field of an observer. At present, reconstructions were made by combining the data of 22 (glaucoma) or 30 or 60 (HH) one-minute trials. As shown in Supplementary Figures S1 and S2, for some movies and simulations, it seems that when using one or two trials, the pattern of the sVFD already emerges.

Somewhat to our surprise, to reconstruct the VFD in the two different simulated pathologies required two rather different approaches. Presumably, this is the case because different types of visual field defects lead to fundamentally different viewing behavior, which we will discuss in detail in the following section.

Different sVFD result in different eye movement behavior and require different approaches for VFD reconstruction

We will now discuss a possible explanation for why the distribution of VP in the visual field is a good indicator for the location of the simulated HH but not for the simulated glaucoma conditions.

In HH, participants need to make many saccades into the masked part of the visual field. If they would not do this, they would end up looking at one side of the screen, with the mask entirely covering it. However, they are not able to direct their gaze towards an interesting part of the scene when saccading into the blind hemifield. Instead, they make large horizontal saccades to swipe away the sVFD (see Figure 9). However, these will be stereotypical and not driven by scene content, explaining the low VP. Making saccades into the blind hemifield is a viewing behavior that is also exhibited by (and taught to) hemianopia patients to improve their visual search performance

(de Haan et al., 2015; Kennard, 2010). This strategy also explains why participants do not make fewer fixations in the complete area of the VFD in the simulated HH conditions than in the control condition. Observers have to direct the same number of fixations to each side of the visual field to not have the screen fully covered by the mask.

However, in the simulated glaucoma conditions, observers may not have an equally predictable viewing strategy to avoid blocking parts of the screen with the sVFD. In the two nasal arc conditions, the mask covers only a small part of the visual field, which makes it easier to predict whether there could be an interesting aspect of the scene covered by the mask. This would be most clearly the case in the small nasal arc condition. Additionally, when the mask covers relevant aspects of the movie scene, participants still have many options for directing their next saccade towards other visible parts of the scene. This may be a lowly prioritized part of the scene, which causes that their next fixation gets assigned a low VP. However, in the case of a small sVFD in particular an observer may still be able to predict, based on information surrounding the masked part, that the relevant part of the scene is hidden by the mask and thus direct their eyes toward it. In that case, the fixation would still be assigned a high VP.

In the simulated tunnel vision condition, the VP is on average lower in the entire visual field than in the other simulated glaucoma conditions, which may not be surprising as a large part of the visual field is masked, which makes it hard to follow the movie clip. Surprisingly, there is a less clear difference in VP between the (visible) center and (masked) periphery of the visual field as there is in the HH condition between the masked and unmasked part of the visual field. This may be due to an enhanced center bias and the short saccades in all directions in this condition.

Participants do not seem to change their eye movement strategy over the time course of the experiment, while they get more used to the presence of the sVFD. At least this does not become apparent in basic eye movement features like fixation duration and saccade amplitude.

Stimuli used to evoke "natural viewing behavior"

Previous studies that investigated how VFDs influence natural viewing behavior have used static images (David et al., 2019) or longer movie clips Crabb et al. (2014). We decided to use movies to have movement as another salient feature that guides viewing behavior. Our movie clips are short, as the content of a movie clip can trigger a specific viewing behavior. By varying the content of our clips, we expected to obtain fixations that were more evenly distributed across the visual field. As Figures 7 and 8 showed, not all movie clips are equally well suited to reconstruct the sVFD. Interestingly, for different archetypes of sVFD, different movie clips seem to be optimal. We conclude that a more variegated set of clips benefits visual field reconstruction.

Artificial scotomas

We simulated the VFD using gray masks to cover parts of the visual field, as is often done (e.g. David et al. (2019). However, this is not how patients experience their VFD. Many patients experience filling-in of the missing parts of the visual field (Crabb et al., 2013). While in static conditions, healthy observers will also often fill in such gray patches, the dynamic conditions and resulting jitter of the simulations in our experiment prevent this. Usually, patients gradually lose luminance sensitivity and this loss is different in both eyes. This means that they can adapt to a slow loss of vision over months and years. Healthy observers, being fully aware of which part of the visual field is masked, may therefore exhibit a different viewing behavior from patients. On the other hand, we know from previous studies that there are systematic differences in viewing behavior between patients and controls, and our methods are aimed at detecting those differences. For example, hemianopia patients make less regular and accurate saccades towards the affected side of their visual field during visual exploration of a scene (Zihl, 1995), which would arguably lead to a lower VP in the affected side of the visual field.

Future work

In future work, our approach needs to be tested with actual patients. Real VFDs are not perceived as gray areas in the field of view, but often are filled in with features from neighboring regions. Therefore we need to test in patients whether our methods are appropriate to detect actual VFDs (Crabb et al., 2013; Hoste, 2003). Moreover, there is still potential to optimize the selection of movie clips, as well as the methods for classification and reconstructing the VFD.

Conclusion

Our findings suggest that it is feasible to use natural viewing behavior recorded while participants view short movie clips to detect the presence of a sVFD. Moreover, it is possible to accurately reconstruct the sVFD. Movie clip viewing in combination with eye tracking may

thus provide an alternative for or complement SAP, in particular for patients unable to perform SAP, such as young children, and vulnerable patients.

Keywords: eye movements, free viewing, homonymous hemianopia, glaucoma, simulated visual field deficit, viewing priority, permutation statistics, support vector machine, perimetry

Acknowledgments

The authors thank Joost Heutink for mediating access to the Psychology student pool.

Supported by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 641805(EGRET) and No. 641805 (NextGenVis) and the Graduate School of Medical Sciences (GSMS), of the University Medical Center Groningen, University of Groningen. The funding organizations had no role in the design, conduct, analysis, or publication of this research.

Presented at the ECVP 2017, ECEM 2017, and Vision2020.

Commercial relationships: none.
Corresponding author: Birte Gestefeld.
Email: b.f.gestefeld@umcg.nl.
Address: Laboratory for Experimental Ophthalmology, University Medical Center Groningen, PO Box 30.001, 9700 RB Groningen, the Netherlands.

References

Asfaw, D. S., Jones, P. R., Smith, N. D., & Crabb, D. P. (2018). Data on eye movements in people with glaucoma and peers with normal vision. *Data in Brief, 19*, 1266–1273, https://doi.org/10.1016/j.dib.2018.05.076.

Bahnemann, M., Hamel, J., De Beukelaer, S., Ohl, S., Kehrer, S., & Audebert, H., …Brandt, S. A. (2015). Compensatory eye and head movements of patients with homonymous hemianopia in the naturalistic setting of a driving simulation. *Journal of Neurology, 262*(2), 316–325, https://doi.org/10.1007/s00415-014-7554-x.

Birt, C. M., Shin, D. H., Samudrala, V., Hughes, B. A., Kim, C., & Lee, D. (1997). Analysis of reliability indices from Humphrey visual field tests in an urban glaucoma population. *Ophthalmology, 104*(7), 1126–1130, https://doi.org/10.1016/S0161-6420(97)30173-0.

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

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers, 34*(4), 613–617, https://doi.org/10.3758/BF03195489.

Cornelissen, Frans W., Bruin, K. J., & Kooijman, A. C. (2005). The influence of artificial scotomas on eye movements during visual search. *Optometry and Vision Science: Official Publication of the American Academy of Optometry, 82*(1), 27–35, https://doi.org/00006324-200501000-00011.

Crabb, D. P., Smith, N. D., Glen, F. C., Burton, R., & Garway-Heath, D. F. (2013). How Does Glaucoma Look? Patient Perception of Visual Field Loss. *OPHTHA, 120*(6), 1120–1126, https://doi.org/10.1016/j.ophtha.2012.11.043.

Crabb, D. P., Smith, N. D., & Zhu, H. (2014). What's on TV? Detecting age-related neurodegenerative eye disease using eye movement scanpaths. *Frontiers in Aging Neuroscience, 6*(NOV), 1–10, https://doi.org/10.3389/fnagi.2014.00312.

David, E. J., Lebranchu, P., Perreira Da Silva, M., & Le Callet, P. (2019). Predicting artificial visual field losses: A gaze-based inference study. *Journal of Vision, 19*(14), 22, https://doi.org/10.1167/19.14.22.

de Haan, G. A., Melis-Dankers, B. J. M., Brouwer, W. H., & Tucha, O. (2015). The Effects of Compensatory Scanning Training on Mobility in Patients with Homonymous Visual Field Defects: A Randomized Controlled Trial, *PLoS One, 10.8*, e0134459, https://doi.org/10.1371/journal.pone.0134459.

Hoste, A. M. (2003). New insights into the subjective perception of visual field defects. *Bulletin de La Société Belge d'ophtalmologie, 287*, 65–71.

Johnson, L. N., & Baloh, F. G. (1991). The accuracy of confrontation visual field test in comparison with automated perimetry. *Journal of the National Medical Association, 83*(10), 895–898.

Kanjee, R., Yücel, Y. H., Steinbach, M. J., González, E. G., & Gupta, N. (2012). Delayed saccadic eye movements in glaucoma. *Eye and Brain, 4*, 63–68, Retrieved from https://www.dovepress.com/getfile.php?fileID=14588.

Kasneci, E., Sippel, K., Heister, M., Aehling, K., Rosenstiel, W., Schiefer, U., . . . Papageorgiou, E. (2014). Homonymous visual field loss and its impact on visual exploration: A supermarket study. *Translational Vision Science and Technology, 3*(6), 2–2. https://doi.org/10.1167/tvst.3.6.2.

Kennard, C. (2010). Compensatory strategies following visual search training in patients with homonymous hemianopia: an eye movement study, 1812–1821, https://doi.org/10.1007/s00415-010-5615-3.

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What's new in Psychtoolbox-3? *Perception, 36*, S14, https://doi.org/10.1068/v070821.

Kübler, T. C., Kasneci, E., Rosenstiel, W., Heister, M., Aehling, K., Nagel, K., … Papageorgiou, E. (2015). Driving with glaucoma: Task performance and gaze movements. *Optometry and Vision Science, 92*(11), 1037–1046, https://doi.org/10.1097/OPX.0000000000000702.

Lamirel, C., Milea, D., Cochereau, I., Duong, M. H., & Lorenceau, J. (2014). Impaired saccadic eye movement in primary open-angle glaucoma. *Journal of Glaucoma, 23*(1), 23–32, https://doi.org/10.1097/IJG.0b013e31825c10dc.

Marsman, J.-B. C., Cornelissen, F. W., Dorr, M., Vig, E., Barth, E., & Renken, R. J. (2016). A novel measure to determine viewing priority and its neural correlates in the human brain. *Journal of Vision, 16*(6), 3, https://doi.org/10.1167/16.6.3.

Skenduli-Bala, E., De Voogd, S., Wolfs, R. C. W., Van Leeuwen, R., Ikram, M. K., & Jonas, J. B., …De Jong, P. T. V. M. (2005). Causes of incident visual field loss in a general elderly population: The Rotterdam study. *Archives of Ophthalmology, 123*(2), 233–238, https://doi.org/10.1001/archopht.123.2.233.

Smith, N. D., Glen, F. C., & Crabb, D. P. (2012). Eye movements during visual search in patients with glaucoma. *BMC Ophthalmology, 12.1*, 1–11.

Tant, M. L. M., Cornelissen, F. W., Kooijman, A. C., & Brouwer, W. H. (2002). Hemianopic visual field defects elicit hemianopic scanning. *Vision Research, 42*(10), 1339–1348, https://doi.org/10.1016/S0042-6989(02)00044-5.

Trauzettel-Klosinski, S., & Brendler, K. (1998). Eye movements in reading with hemianopic field defects: The significance of clinical parameters. *Graefe's Archive for Clinical and Experimental Ophthalmology, 236*(2), 91–102, https://doi.org/10.1007/s004170050048.

Wild, J. M. (1988). Techniques and developments in automated perimetry: A review. *Ophthalmic and Physiological Optics, 8*(3), 295–308, https://doi.org/10.1111/j.1475-1313.1988.tb01059.x.

Zihl, J. (1995). Visual scanning behavior in patients with homonymous hemianopia. *Neuropsychologia, 33*(3), 287–303, https://doi.org/10.1016/0028-3932(94)00119-A.