Objective Estimation of Visual Acuity with Preferential Looking

Veit Sturm,[1,2] *Daniel Cassel,*[3,4] *and Moshe Eizenman*[1,3,4]

Purpose. A novel preferential looking (PL) procedure that uses quantitative analysis of visual scanning parameters is presented.

Methods. Nine adult subjects were presented with a set of 14 visual stimuli (stimuli included three uniform gray fields and one field with black-and-white square wave gratings) spanning the range of spatial frequencies from 1.5 cyc/deg to 35.1 cyc/deg (1.3 logMAR to −0.07 logMAR). A remote gaze-tracking system was used to monitor the subject's eye movements and the relative fixation time (RFT) on the grating target. Subsequently, a four alternative forced-choice psychophysical test (4AFC) was performed with the same visual stimuli.

Results. For visual stimuli for which the gratings' positions in the 4AFC test were identified correctly in 100% of the trials (reliably discriminated), the mean RFT was 72.5% ± 9.0%. For stimuli for which the spatial frequencies were higher than the subject's psychophysically determined visual acuity (VA) threshold (nondiscriminated), the mean RFT was 25.3% ± 8.5%. Using three repeated trials at each spatial frequency and a VA detector based on the conditional probability density functions of the RFT, the average VA was underestimated by 0.06 logMAR (range, 0.00–0.20 logMAR).

Conclusions. In adults, automated quantitative analysis of visual scanning patterns can be used to estimate VA objectively and rapidly (210 seconds) with a mean error of 0.06 logMAR. The novel approach may form the basis for PL procedures that are more objective and more accurate than the traditional clinical PL procedures. (*Invest Ophthalmol Vis Sci.* 2011;52:708–713) DOI:10.1167/iovs.09-4911

Human infants prefer to fixate patterned surfaces more than homogeneous ones and respond to visual stimuli by moving their eyes in the direction of the object of visual interest.[1,2] This natural fixation and tracking behavior was monitored in the visual preference method of Fantz.[3] The duration of the fixation was measured by adult observers of each of the two stimuli, and the mean percentage of fixation time was calculated.[3,4] The stimulus with significantly longer percentage of fixation time was taken as the "preferred stimulus." Infant preferences were not derived; rather, the data were averaged for groups of infants.[3,4]

In forced-choice preferential looking (FPL), Teller et al.[5] combined the visual preference approach with the two-alternative forced-choice psychophysical method.

The FPL technique was conceived as an objective measurement of visual acuity (VA) thresholds in infants. Subsequently, various research groups have applied the FPL approach (or its operant modification) to derive psychometric functions for diverse visual parameters, such as VA, stereopsis, color vision, contrast sensitivity, and dark and light adaptation.[6–11] In FPL, the observer is masked to the stimulus location and judges the stimulus side based on the infant's looking behavior (combined eye and head movements). The observer is provided with trial-by-trial feedback regarding whether the judgment is correct. Thus, the infant's behavior is directly related to the critical stimulus parameter. A criterion performance level—the highest spatial frequency for which 75% of the responses in judging the location of the target with the gratings are correct—is then used as a measurement of the infant's VA.[5,12,13] A reliable threshold estimation of VA by the FPL technique required multiple repeated trials for each level of spatial frequency (approximately 20–25).[5,12,13]

Because the length of the test (15–45 minutes) limited the clinical usefulness of the technique, several modifications have been proposed. Variants such as an age-dependent diagnostic stripe width[12,14–16] and the up-down staircase method of stimulus presentation[17] can be regarded as steps into daily clinical routine. However, the procedures tended to be more variable than the laboratory procedures and still took more time than was feasible for wide clinical application.[14–17]

Eventually, introduction into clinical practice was achieved with the acuity card procedure (ACP).[18] Instead of testing a fixed number of presentations in a fixed protocol, this methodology uses an experienced observer's subjective integrated judgment regarding the infant's looking behavior. The procedure shortened testing time and improved testing success in infants considerably.[18] In addition, both the ACP and the FPL techniques had similar variability.[6,18–21] Hence, Teller acuity cards (TACs) have been established as the standard diagnostic tool for evaluating VA in infants and preverbal children.[22]

The purpose of this article is to present a novel PL technique that can be both statistically stable and efficient. The technique is based on the analysis of visual scanning patterns that are monitored by a gaze estimation system and the use of an efficient stimulus presentation method. The performance of the novel technique is evaluated in a study with adults.

From the [1]Department of Ophthalmology and Visual Sciences, The Hospital for Sick Children and the University of Toronto, Toronto, Ontario, Canada; the [2]Department of Ophthalmology, University Hospital of Zurich, Zurich, Switzerland; and the [3]Institute of Biomaterials and Biomedical Engineering and the [4]Department of Electrical Engineering and Computer Science, University of Toronto, Toronto, Ontario, Canada.

Supported by the Vision Science Program at the Toronto Western Hospital, the Natural Sciences and Engineering Research Council of Canada, The Canadian Institute for Health Research, and the Nachwuchsförderungskredit Stiefel-Zangger-Stiftung of the University of Zurich, Switzerland.

Submitted for publication November 12, 2009; revised May 13 and July 24, 2010; accepted July 28, 2010.

Disclosure: **V. Sturm**, None; **D. Cassel**, None; **M. Eizenman**, None

Corresponding author: Veit Sturm, Department of Ophthalmology, University Hospital of Zurich, Frauenklinikstrasse 24, CH-8091 Zurich, Switzerland; veit_sturm@yahoo.com.

Subjects, Materials, and Methods

Subjects

Nine naive adult subjects (four women, five men) were recruited from students at the University of Toronto, Canada. Subjects' ages ranged from 23 to 35 years (mean, 25.9 years). This study adhered to the tenets of the Declaration of Helsinki. The complete protocol was reviewed and approved by the Ethics Committee at the University of Toronto before commencement, and all patients gave informed consent before participating in the study. Subjects with a history of ophthalmic deficits other than refractive error were excluded. An ophthalmic examination including best-corrected VA measurement (20/20 or better in all cases), slit-lamp assessment of the anterior segment, and direct ophthalmoscopy was performed. The mean spherical equivalent refractive error of the tested eyes was −3.1 diopters (range, 0 to −6.5 diopters).

Experimental Setup and Visual Stimuli

Visual stimuli were presented on a 21-inch CRT computer monitor. Each visual stimulus consisted of a 2 × 2 grid with three uniform gray homogeneous fields and one field with black-and-white square wave gratings (Fig. 1). Each field was presented as a rectangular target easily distinguishable from the black background. All fields were the same size (8.2° × 6.15°), and the center of each field was 7.5° from the center of the screen (at a viewing distance of 83 cm). The separation between two adjacent targets was 4.6 cm (3.2°). Luminance was measured with a luminance meter (model LS-100; Minolta, Osaka, Japan). The black-white Michelson contrast of the gratings ranged from 96% to 97%. The luminance of the black stripes was 6.75 cd/m^2 (i.e., similar to the background), whereas the luminance of the white stripes, which was measured at three different positions, varied from 384 cd/m^2 to 405 cd/m^2 (average, 393 cd/m^2; 2.6 log cd/m^2). The luminance of the gray fields varied from a minimum of 190 cd/m^2 to a maximum of 210 cd/m^2 (average, 200 cd/m^2). The maximum difference between the space-average luminance of the three gray fields in each visual stimulus was <5%. For each visual stimulus, the space-average luminance of the field with the gratings was within 2% of the space-average luminance of the average of the three gray fields. Given that the differences in space average luminance among the plain fields were similar to or larger than the differences in space average luminance between the plain fields and the field with the gratings, subjects could not use luminance as a cue to determine the field with the gratings. The field with the gratings could appear randomly at 1 of 4 positions.

The 14 spatial frequencies used were 1.5, 2.3, 3.12, 4.68, 6.24, 9.36, 12.48, 14.62, 18.73, 21.06, 24.96, 29.25, 32.76, and 35.1 cyc/deg at a distance of 83 cm. This large range of frequencies was used to build the probability density functions (see Results). The range of spatial frequencies was similar to the range of spatial frequencies tested by the TAC at 55 cm (0.31, 0.42, 0.63 0.84, 1.30, 1.60, 2.40, 3.10, 4.70, 6.40, 9.60, 13.00, 19.00, 26.00, 38.00) with higher spatial frequencies added (21.06, 29.25, 32.76) and lower spatial frequencies deleted (0.31, 0.42, 0.63, 0.84). This was done to improve the resolution of the measurements when testing adults because their VA would lie in the higher spatial frequencies whereas the range provided by the TAC is sparse at these higher spatial frequencies. For graphic purposes, the spatial frequencies were converted from cycles per degree to the log of the minimum angle of resolution (logMAR). The minimum angle of resolution is expressed in minutes of arc (1 min arc = 1/60th of a degree); thus, logMAR = log (60/spatial frequency of gratings * 2).

A remote gaze-tracking system was used to monitor and estimate the subject's visual scanning parameters (Vision 2020; El-Mar Inc., Toronto, Canada). The gaze estimation system extracts eye features from video images and uses these features to estimate gaze position. The features extracted are the pupil center and two or more corneal reflexes. The corneal reflexes are virtual images of infrared light sources illuminating the eye and are created by the front surface of the cornea. The gaze-tracking system allows free head movements in a volume of 25 * 25 * 25 cm^3 and estimates gaze position at a rate of 30 Hz with accuracy better than 1°.[23–25] The system was optimized to measure the point-of-gaze in infants and young children (i.e., calibration routine requires only one point).

The relative fixation time (RFT), which is defined as the percentage of fixation time on the field with the gratings over the sum of fixation times on all four fields, was used to quantify visual scanning behavior.

Experimental Procedure

Thresholds for grating detection were obtained using two different methods. Gaze tracking was always performed first because subjects were required to be completely naive to the stimuli and was followed by psychophysical testing. One eye (subject's preference) of each subject was patched, and the subject was seated 83 cm from the computer monitor. To obtain a broader range of VA thresholds, six myopic participants were tested without optical correction.

Subjects were instructed to look at a point on the computer monitor for 3 seconds to complete a 1-point calibration procedure on the gaze-estimation system.[24] Participants were instructed at the beginning of the gaze-tracking experiment to simply look toward the computer monitor. The stimuli were presented from the lowest spatial frequency to the highest spatial frequency in ascending order. Each spatial frequency was tested three times, and each test lasted 5 seconds. There was no delay between the three tests, but for each test the field with the gratings appeared in a different location on the computer monitor. The total time for each gaze-tracking experiment was 210 seconds.

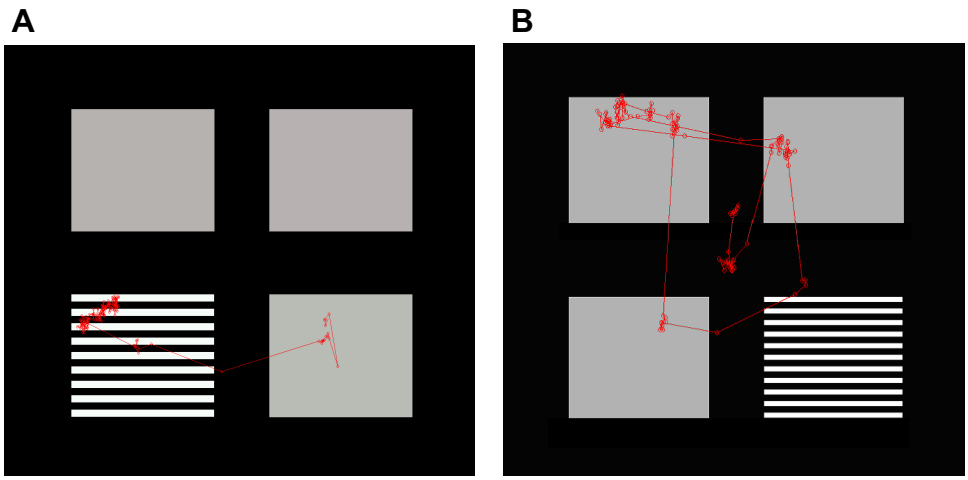

Figure 1. Visual scanning patterns. Gaze position (*circles*) for one trial (5 seconds) superimposed on a schematic description of the visual stimulus. (**A**) Scanning pattern when the spatial frequency of the gratings was lower than the subject's psychophysical VA. (**B**) Scanning pattern when the spatial frequency of the gratings was higher than the subject's psychophysical VA.

Subsequently, each subject completed a four-alternative forced-choice (4AFC) psychophysical test. The same stimulus display and the same test stimuli were used. Stimuli were presented with spatial frequencies in ascending order. Subjects were instructed to identify the target with the gratings by providing a verbal response. No feedback was given. Each spatial frequency was tested 10 times. VA threshold was set at a point where the subject correctly detected the location of the field with the gratings in more than six of the 10 trials (performance level approximately halfway between chance level 25% and 100%).

Results

Analysis of Visual Scanning Patterns

Figure 1 shows examples of two different visual scanning patterns. Figure 1A shows eye movements of a subject when the spatial frequency of the grating was lower than the subject's psychophysical VA. When the visual stimulus appeared, the subject was fixating on the bottom-right field. After a short delay (<400 ms), the subject initiated a saccade toward the grating and kept fixating on the grating for the rest of the trial. Figure 1B shows eye movements of a subject when the spatial frequency of the grating was higher than the subject's psychophysical VA. In this case, the subject scanned all four fields but spent more time fixating on one of the homogeneous fields than on the field with the grating.

Figure 2A presents the psychometric function of the RFT for one subject (subject 4). The average RFT, for all nine subjects, as a function of the distance from the psychophysical threshold for each subject is shown in Figure 2B. Here the RFT psychometric functions of the subjects were aligned (shifted along the *x*-axis) so that for each subject 0 logMAR was set at the subject's psychophysically determined VA threshold. Each point in Figure 2B is the average RFT for all subjects over a range of $x \pm 0.05$ logMAR. For example, the data presented in Figure 2B at $x = 0.1$ logMAR from threshold are the average RFTs for all subjects for stimuli that are 0.05 logMAR to 0.15 logMAR from each subject's psychophysically determined VA threshold. Based on the results of the 4AFC psychophysical tests, one can define two different regions for the RFT in Figure 2B. The first region includes all spatial frequencies for which subjects could discriminate between the field with the grating and the homogeneous fields in 100% of the psychophysical tests (reliably discriminated), and the second includes all spatial frequencies that were higher than the subjects' psychophysically determined VA thresholds (nondiscriminated). The mean RFT for the reliably discriminated region was 72.5% ± 9.0%, and the mean RFT for the nondiscriminated region was 25.3% ± 8.5%.

Determination of Visual Acuity

The probability density functions of the RFT in the reliably discriminated and the nondiscriminated regions are used by an optimal detector (likelihood ratio test; see Appendix) to determine whether the null hypothesis, H_0 (subject cannot discriminate between the field with the gratings and the homogeneous fields), or the alternative hypothesis, H_1 (subject can discriminate between the field with the gratings and the homogeneous fields), should be accepted. Based on the results of the likelihood ratio test at each spatial frequency, the subject's VA is determined as the highest spatial frequency (lowest logMAR) for which the probability of false acceptance of H_1 is smaller than the probability of misdetection of gratings with lower spatial frequencies.

The following examples demonstrate the estimation of VA for two of the subjects in the study. The examples are based on three trials at each spatial frequency. As is shown in the Appendix, for a probability of false positive, P_F (false acceptance of H_1), of 5% at each spatial frequency, the probability of detection, P_D, is 88.6%, and the probability of misdetection, $P_M = 1 - P_D$, is 11.4%. The results of the likelihood ratio test for all subjects are summarized in Table 1.

For subject 3, for example, the highest spatial frequency (lowest logMAR) for which H_1 was accepted was 0.15 logMAR. Given that for all gratings with lower spatial frequencies H_1 was accepted, 0.15 logMAR was accepted as the VA estimate for subject 3. On the other hand, for subject 5, the highest spatial frequency for which H_1 was accepted was 0.21 logMAR. Because H_1 was rejected for two gratings with lower spatial frequencies (0.30 logMAR and 0.38 logMAR), the probability of misdetection was $P_M^2 = 0.114^2 = 0.013$. Given that the probability of false acceptance of H_1 at 0.21 logMAR was 5%, which was larger than the probability of misdetection at the two lower spatial frequencies (1.3%), 0.21 logMAR was not accepted as the VA estimate for subject 5. When the highest spatial frequency for which H_1 was accepted by the likelihood ratio test was rejected as a VA estimate, the process was repeated for the second highest spatial frequency for which H_1 was accepted (i.e., 0.50 logMAR). For subject 5, 0.50 logMAR was accepted as the VA.

Using this procedure to estimate VA, the average VA for all subjects was underestimated by 0.11 logMAR (range, −0.73 to 0.37 logMAR) when one trial was used at each spatial frequency (the subject's psychophysically determined VA threshold was used as the "true" VA). Average VA was underestimated by 0.11 logMAR (range, −0.38 to 0.17 logMAR) when two trials were used at each spatial frequency, and average VA was underestimated by 0.06 logMAR (range, 0.00 to 0.20

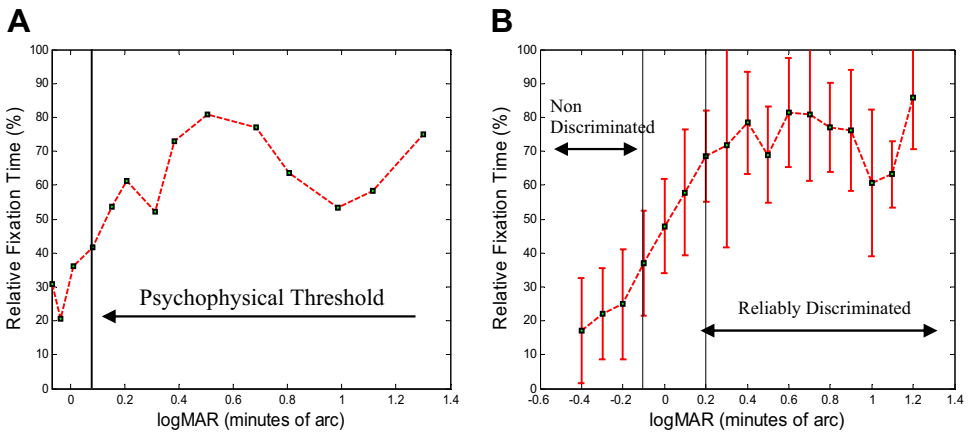

Figure 2. (**A**) Psychometric function of the relative fixation time for subject 4. Each data point represents the average of three trials. (**B**) Average RFT for all subjects as a function of distance from the psychophysical threshold (0 logMAR). Error bars represent SD.

TABLE 1. Results of the Likelihood Ratio Tests

	LogMAR (min arc)													
Subject	**1.3**	**1.12**	**0.99**	**0.81**	**0.68**	**0.50**	**0.38**	**0.30**	**0.21**	**0.15**	**0.08**	**0.01**	**−0.04**	**−0.07**
1	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y*	N	N	N
2	Y	Y	Y	Y	Y	Y	Y*	N	N	N	N	N	N	N
3	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y*	N	N	N	N
4	Y	Y	Y	Y	Y	Y	Y	N*	N	N	N	N	N	N
5	Y	Y	Y	Y	Y	Y	N*	N	Y	N	N	N	N	N
6	Y	Y	Y	Y	Y	Y	N	N*	N	N	N	N	N	N
7	Y	Y	Y	Y	Y	Y	N	Y	N*	N	N	N	N	N
8	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	N*	N	N	N
9	N	N	Y	N	Y	Y	Y	Y	Y	Y*	N	N	N	N

Y, accept hypothesis H_1; N, accept hypothesis H_0.
* Psychophysically determined VA threshold.

logMAR) when three trials were used at each spatial frequency.

DISCUSSION

The original quantitative visual preference technique that was developed by Fantz[1–3] to determine infants' visual capabilities evolved and changed over the years into a laboratory test (FPL) and a clinical test (ACP). In the past two decades, variants of ACP using printed gratings, such as TACs (TAC I [Vistech, Inc., Dayton, OH] and TAC II [Stereo Optical Co., Inc., Chicago, IL]) have become the standard diagnostic tool for evaluating VA in infants and nonverbal children. In the ACP with TAC grating stimuli, an examiner presents the stimuli on acuity cards to the infant and makes a judgment regarding the infant's ability to detect the gratings. Although the examiner remains objective regarding the right/left position of the grating, the integrative subjective judgment of grating threshold is based on qualitative assessment of the infant's eye movements, head movements, facial expressions, pointing, and verbal responses.[18,26–28] Although this integrative subjective approach may help the examiner make a decision, none of the parameters is quantitatively assessed, and there is no form of outside control over the objectivity of the examiner.

Because VA estimations obtained with preferential looking procedures depend on subjective judgements, intraobserver and interobserver reliability of the results are important for data interpretation. Most previous studies dealt with interobserver reliability. Examination was performed in different populations across a wide range of ages, including those with normal vision and severely disabled children. Good interobserver agreement for ACP (≤1 octave [0.3 logMAR] difference of test-retest scores) has been found in 86% to 98% of children with normal vision.[29–31] Children with ocular or neurologic abnormalities showed somewhat poorer interobserver agreement with ACP.[32–34] Interobserver test-retest reliability is similar for ACP and the more scientifically rigorous, laboratory-based FPL.[6,18–21,31,34] Only a few studies have examined intraobserver test-retest reliability in ACP. The test-retest pairs differed by no more than 1 octave for 88% to 99% of healthy full-term infants.[18,30] Similar results (91% of intraobserver test-retest scores differed by no more than 1 octave) were obtained in children who were born before term or who had perinatal complications, or both.[35] On the other hand, intraobserver agreement was considerably lower when severely disabled children (Down syndrome, mental retardation, cerebral palsy) were tested. The test-retest pairs of VA estimates differed by 1 octave or more for 17% to 53% of the subjects.[36–38] In general we attribute this variability to inherent subject variability (noise derived from the infants' receptors and their detection and response mechanisms) and observer's factors (integrative, subjective judgment about infants' responses).

The PL technique herein described eliminates the latter. In other words, it removes the noise in FPL or ACP contributed by the observer but does not affect the noise contributed by the infant. To overcome the tester's necessary intervening judgments in FPL and ACP, the novel PL technique presented in this article uses the statistical properties of the RFT, a parameter that describes the subject's visual scanning behavior. The automated detector uses conditional probability density functions of the RFT to determine changes in visual scanning behavior when subjects can and cannot detect the target with the gratings.

As shown in the Appendix, the accuracy and reliability of the technique increases with the number of trials at each spatial frequency. In adults, VA in a range from 1.3 logMAR to −0.07 logMAR can be estimated objectively and efficiently (in 210 seconds) with an average error of 0.06 logMAR (range, 0.00–0.20 logMAR). Although increasing the number of trials does increase the accuracy (i.e., the probability of detection), the increase in accuracy decreases as more trials are used. For instance, when three trials are used, the probability of detection is 88.6%. Using four trials increases the probability of detection to 94.7%, and increasing the number of trials to five results in a probability of detection of 97.6%. Meanwhile, as the number of trials is increased, the testing time is increasing linearly. Thus the tradeoff between processor accuracy and testing time becomes more and more undesirable as more trials are used. Hence, the results with three trials provide a reasonable balance between accuracy and efficiency (i.e., testing time). Efficiency can be further improved by limiting the range over which the VA is estimated for each subject (similar to current clinical practice with the ACP) and by using adaptive presentation strategies such as the up-down staircase method of stimulus presentation[17] or the combined fast "screening" and fine threshold testing protocol.[39] When compared with the assessment of VA by the ACP or FPL techniques, the novel PL-fixation technique is much more efficient (by almost 1 order of magnitude).

The present study is the first step in the development of a rapid, statistically robust PL technique for VA assessment in infants. However, in this context, several issues must be considered. To use the novel PL technique with infants and young children one has to be able to use visual stimuli that provide more than two alternative positions for the gratings. Furthermore, the VA detector has to be constructed with conditional probability density functions of the RFT that are suitable for infants and young children. Teller[13] explored the possibility of

using three or four alternative grating positions to increase the efficiency of the test. These two presentation scenarios generally worked with infants, though the expected increase in efficiency (lower probability of false positive) was not realized because many trials had to be rejected as a result of the observers' difficulties in deciding which of the fields was fixated preferentially. Possible general difficulties of infants making clearly preferential eye movements with four alternatives might have further impeded the observers' decision-making. Using an accurate eye-tracking system to determine infant's fixation should improve the ability to determine which of the fields is fixated preferentially when larger numbers of alternatives are present. The main challenge with infants is to keep the infant's attention throughout the testing. To keep the infant's attention, it might be necessary to significantly increase the size of the monitor so that infants will be less distracted by the surroundings and to integrate an operant approach of rewarding "proper" visual scanning behavior for above-threshold stimuli (e.g., triggering cartoons when the infant is looking at the gratings).

In conclusion, this study demonstrates the value of quantitative analysis of visual scanning patterns in a PL procedure with adults. The use of an accurate eye-tracking system to monitor eye movements supports efficient presentation of visual stimuli and enables the construction of efficient and accurate VA estimators. It is important to emphasize that even though the VA detector in this study was based on the statistical properties of a single parameter (RFT), in principle the VA detector can be constructed with the joint probability density functions of multiple visual scanning parameters (e.g., probability of first fixation on the field with the gratings). Such an extension might further improve the performance of the detector (assuming that not all the visual scanning parameters are perfectly correlated). On the other hand, some of the other parameters that are used to judge the infant's ability to detect gratings (e.g., eyes' widening) might be lost using a gaze-tracking device.

The new approach to the analysis of visual scanning parameters can provide the basis for novel PL procedures that may have greater accuracy than the traditional ACP.

Appendix

The conditional probability density functions of the RFT for spatial frequencies in the reliably discriminated region, $f(RFT|H_1)$, and in the nondiscriminated region, $f(RFT|H_0)$, are presented in Figure A1. The probability density functions are based on data from all nine subjects (251 observations in the reliably discriminated region and 109 observations in the nondiscriminated region).

The two conditional probability density functions of the RFT were used to construct an optimal test to determine whether the null hypothesis, H_0 (subject cannot discriminate between the target with the grating and the homogeneous targets), or the alternative hypothesis, H_1 (subject can discriminate between the target with the grating and the homogeneous targets) increase, should be accepted.[40] The optimality criterion for the test was to maximize the probability of detection, P_D (accepts H_1 when H_1 is true), for a fixed probability of false positive, P_F (accepts H_1 when H_0 is true). Under this criterion, the likelihood ratio test is optimal.[41] The likelihood ratio for the parameter RFT, for N independent trials ($1 \ldots N$), can be written as the ratio of the products of the conditional probability density functions in the reliably discriminated region, $f(RFT_i|H_1)$ (Fig. A1A), over the conditional probability density functions in the nondiscriminated region, $f(RFT_i|H_0)$ (Fig. A1B):

$$\Lambda(RFT_1, RFT_2, \ldots RFT_N) = \frac{f(RFT_1|H_1) * f(RFT_2|H_1) * \ldots f(RFT_N|H_1)}{f(RFT_1|H_0) * f(RFT_2|H_0) * \ldots f(RFT_N|H_0)} \quad (1)$$

If the likelihood ratio, $\Lambda(RFT_1, RFT_2, \ldots RFT_N)$, is $\leq \lambda$, accept H_0; otherwise, accept H_1.

Using the probability density functions in Figure A1, the conditional probability density of the likelihood ratio when H_0 is true, $f(\Lambda|H_0)$, and when H_1 is true, $f(\Lambda|H_1)$, can be computed. The probability of false alarm, P_F, (accepts H_1 when H_0 is true), and the probability of detection, P_D, (accepts H_1 when H_1 is true) can then be calculated from equations 2 and 3, respectively. The value of λ is determined by setting P_F to a certain level (e.g., 5%)

$$P_F = \int_{\lambda}^{\infty} f(\Lambda \mid H_0) d\Lambda \quad (2)$$

$$P_D = \int_{\lambda}^{\infty} f(\Lambda \mid H_1) d\Lambda \quad (3)$$

when P_F is set to 5% ($\alpha = 5\%$) and only one trial is used at each spatial frequency, $\lambda = 3.63$ and $P_D = 40\%$. When two trials are used at each spatial frequency, for $\alpha = 5\%$, $\lambda = 2.51$ and $P_D = 75.1\%$. If data from three trials are used, $\alpha = 5\%$, $\lambda = 1.58$, and $P_D = 88.6\%$.

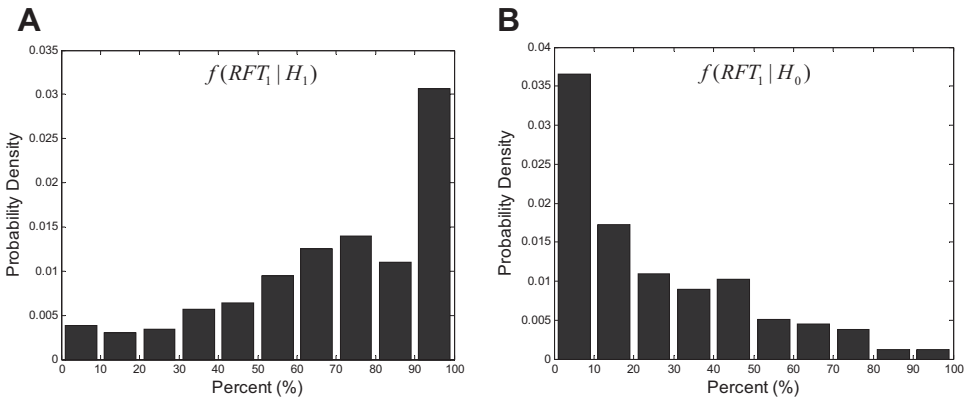

Figure A1. Conditional probability density functions for the relative fixation time. (**A**) Grating frequencies in the reliably discriminated region. (**B**) Grating frequencies in the nondiscriminated region.

Acknowledgments

The authors thank Elias Guestrin for helping to develop the technology and the anonymous reviewers for their truly helpful comments.

References

1. Fantz RL. Pattern vision in young infants. *Psychological Records.* 1958;8:43–48.
2. Fantz RL, Ordy J, Udelf M. Maturation of pattern vision in infants during the first 6 months of life. *J Comp Physiol Psychol.* 1962; 55:907–917.
3. Fantz RL. Visul perception from birth as shown by pattern selectivity. *Ann N Y Acad Sci.* 1965;118:793–814.
4. Fantz RL. Pattern vision in newborn infants. *Science.* 1963;140: 296–297.
5. Teller DY, Morse R, Borton R, Regal D. Visual acuity for vertical and horizontal gratings in human infants. *Vision Res.* 1974;14: 1433–1439.
6. Gwiazda J, Brill S, Mohindra I, Held R. Infant visual acuity and its meridional variation. *Vision Res.* 1978;18:1557–1564.
7. Banks M, Salapatek P. Acuity and contrast sensitivity in 1-, 2-, and 3-month-old human infants. *Invest Ophthalmol Vis Sci.* 1978;17: 361–365.
8. Held R, Gwiazda J, Brill S, Mohindra I, Wolfe J. Infant visual acuity is underestimated because near threshold gratings are not preferentially fixated. *Vision Res.* 1979;19:1377–1379.
9. Birch EE, Gwiazda J, Bauer JA Jr, Naegele J, Held R. Visual acuity and its meridional variations in children aged 7–60 months. *Vision Res.* 1983;23:1019–1024.
10. Atkinson J, Braddick O. Assessment of visual acuity in infancy and early childhood. *Acta Ophthalmol Suppl.* 1983;157:18–26.
11. Birch EE. Infant interocular acuity differences and binocular vision. *Vision Res.* 1985;25:571–576.
12. Dobson V, Teller DY, Ping Lee C, Wade B. A behavioural method for efficient screening of visual acuity in young infants, I: preliminary laboratory development. *Invest Ophthalmol Vis Sci.* 1978; 17:1142–1150.
13. Teller DY. The forced-choice preferential looking procedure: a psychophysical technique for use with human infants. *Inf Behav Dev.* 1979;2:135–153.
14. Fulton AB, Manning KA, Dobson V. A behavioural method for efficient screening of visual acuity in young infants, II: clinical application. *Invest Ophthalmol Vis Sci.* 1978;17:1151–1157.
15. Gwiazda J, Wolfe JM, Brill S, Mohindra I, Held R. Quick assessment of preferential looking acuity in infants. *Am J Optom Physiol Opt.* 1980;57:420–427.
16. Gwiazda J, Brill S, Mohindra I, Held R. Preferential looking acuity in infants from two to fifty-eight weeks of age. *Am J Optom Physiol Opt.* 1980;57:428–432.
17. Mayer DL, Fulton AB, Hansen RM. Preferential looking acuity obtained with a staircase procedure in pediatric patients. *Invest Ophthalmol Vis Sci.* 1982;23:538–543.
18. McDonald MA, Dobson V, Sebris SL, Baitch L, Varner D, Teller DY. The acuity card procedure: a rapid test of infant acuity. *Invest Ophthalmol Vis Sci.* 1985;26:1158–1162.
19. Mayer DL, Dobson V. Visual acuity development in infants and young children, as assessed by operant preferential looking. *Vision Res.* 1982;22:1141–1151.
20. Atkinson J, Braddick O, Pimm-Smith E. Preferential looking for monocular and binocular acuity testing of infant. *Br J Ophthalmol.* 1982;66:264–268.
21. Dobson V, Salem D, Carson JB. Visual acuity in infants: the effects of variations in stimulus luminance within the photopic range. *Invest Ophthalmol Vis Sci.* 1983;24:519–522.
22. Drover JR, Wyatt LM, Stager DR, Birch EE. The Teller Acuity Cards are effective in detecting amblyopia. *Optom Vis Sci.* 2009;86:755–759.
23. Guestrin ED, Eizenman M. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Trans Biomed Eng.* 2006;53:1124–1133.
24. Guestrin ED, Eizenman M. Remote point-of-gaze estimation requiring a single-point calibration for applications with infants. Savannah, GA: *Proceedings of the 2008 Symposium on Eye Tracking Research and Applications*; ACM:267–274.
25. Model D, Eizenman M, Sturm V. Fixation-free assessment of the hirschberg ratio. *Invest Ophthalmol Vis Sci.* 2010;51:4035–4039.
26. Courage ML, Adams RJ. Visual acuity assessment from birth to three years using the acuity card procedure: cross-sectional and longitudinal samples. *Optom Vis Sci.* 1990;67:713–718.
27. Mayer DL, Beiser AS, Warner AF, Pratt EM, Raye KN, Lang JM. Monocular acuity norms for the Teller Acuity Cards between ages one month and four years. *Invest Ophthalmol Vis Sci.* 1995;36: 671–85.
28. Johnson C, Kran BS, Deng L, Mayer DL. Teller II and Cardiff Acuity testing in a school-age deafblind population. *Optom Vis Sci.* 2009; 86:188–195.
29. McDonald MA, Sebris SL, Mohn G, Teller DY, Dobson V. Monocular acuity in normal infants: the Acuity Card procedure. *Am J Optom Physiol Opt.* 1986;63:127–134.
30. Mayer DL, Beiser AS, Warner AF, Pratt EM, Raye KN, Lang JM. Monocular acuity norms fort he Teller Acuity Cards between ages one month and four years. *Invest Ophthalmol Vis Sci.* 1995;36: 671–685.
31. Getz LM, Dobson V, Luna B, Mash C. Interobserver reliability of the teller acuity card procedure in pediatric patients. *Invest Ophthalmol Vis Sci.* 1996;37:180–187.
32. Preston KL, McDonald MA, Sebris SL, Dobson V, Teller DY. Validation of the Acuity Card procedure for assessment of infants with ocular disorders. *Ophthalmology.* 1987;94:644–653.
33. Dobson V, Carpenter NA, Bonvalot K, Bossler J. The acuity card procedure: interobserver agreement in infants with perinatal complications. *Clin Vision Sci.* 1990;6:39–48.
34. Mash C, Dobson V, Carpenter N. Interobserver agreement for measurement of grating acuity and interocular acuity differences with the Teller Acuity Card procedure. *Vision Res.* 1995;35:303–312.
35. Mash C, Dobson V. Intraobserver reliability of the Teller Acuity Card procedure in infants with perinatal complications. *Optom Vis Sci.* 2005;82:817–822.
36. Hertz BG. Acuity card testing of retarded children. *Behav Brain Res.* 1987;24:85–92.
37. Hertz BG, Rosenberg J. Acuity card testing of spastic children: preliminary results. *J Pediatr Ophthalmol Strabismus.* 1988;25: 139–144.
38. Hertz BG. Use of the acuity card method to test retarded children in special schools. *Child Care Health Dev.* 1988;14:189–198.
39. Holmes JM, Beck RW, Repka XM, et al. The amblyopia treatment study visual acuity testing protocol. *Arch Ophthalmol.* 2001;119: 1345–1353.
40. Cassel D. *An Objective Methodology to Assess Visual Acuity Using Visual Scanning Parameters* (master's thesis). Toronto, Ontario, Canada: University of Toronto; 2009.
41. Van Trees HL. *Detection, Estimation and Modulation Theory*. New York: John Wiley & Sons Inc.; 1968.