Comparison of image-based quantification methods in evaluating fixation stability using a remote eye tracker in abnormal phoria

Sang-Yeob Kim, Byeong-Yeon Moon, Hyun Gug Cho and Dong-Sik Yu [ORCID]

Abstract

Objective: This study was performed to establish a quantitative evaluation and comparison of fixation stability, as measured by an eye tracker, using image-based areas determined by the bivariate contour ellipse area (BCEA), kernel density estimation (KDE), and Scanpath methods.

Methods: This prospective cross-sectional study included 45 and 20 participants with abnormal and normal phoria, respectively. Eye movements were recorded using a remote eye tracker and were plotted using RStudio software. Image-based areas were evaluated using ImageJ software.

Results: The image-based areas used to evaluate fixation stability exhibited decreasing stability in the abnormal phoria group in the following order: KDE with ±1 standard deviation (SD), BCEA with ±1 SD, KDE with ±2 SD or Scanpath, and BCEA with ±2 SD. The BCEA tended to be overestimated, and the KDE tended to be underestimated at high density. The Scanpath method had a very high probability area because the area spans all gaze points.

Conclusions: Fixation stability could be quantified as image-based areas by the KDE, BCEA, and Scanpath methods. Our findings suggest that fixation stability may be evaluated using one or more methods.

Keywords

Fixation stability, bivariate contour ellipse area, kernel density estimation, scanpath, binocular vision, eye tracker

Department of Optometry, Kangwon National University, Samcheok, South Korea

Corresponding author:
Dong-Sik Yu, Hwangjogil 346, Dogye-up, Samcheok, Gangwondo 25949, Republic of Korea.
Email: yds@kangwon.ac.kr

Introduction

Fixation stability evaluates the ability of the eyes to maintain steady fixation on a target. Fixation disorders can indicate organic or functional anomalies, such as nystagmus, strabismus, amblyopia, cataract, glaucoma, maculopathy, and ocular myasthenia gravis.[1–5] Clinical evaluation of fixation stability may involve direct observations by clinicians and the objective recording of eye movements using electrooculographic instruments, such as the Readalyzer and Visagraph II.[6,7] The use of an eye tracker to evaluate visual functions, including eye fixation and movements, is increasing. Modern eye-tracking systems can measure fixation stability during binocular viewing under natural viewing conditions and can be worn with spectacles or contact lenses.[8,9]

The quantification of fixation stability using an eye tracker can be performed using several methods: standard deviation (SD) of eye gaze positions from the mean eye gaze;[10] mean Euclidean distance, which is a measure of the central dispersion of fixation positions around the mean position of gaze points;[11] bivariate contour ellipse area (BCEA) of horizontal and vertical components;[12,13] and kernel density estimation (KDE) of the probability density function.[14,15] The BCEA is considered the gold standard measure of fixation stability.[16] As per previous studies, although the BCEA is the most commonly used metric, the KDE describes fixation more fully than the BCEA in macular disease.[15] Euclidean distance, based on sequential fixation locations, is better than the BCEA for evaluating fixation stability in glaucoma.[17] Fixation instability can also be a useful tool for mass screening children to diagnose strabismus in the absence of amblyopia and latent nystagmus.[18]

In studies, comparisons between current tools using methods such as the BCEA and KDE might be inadequate to evaluate and characterize fixation patterns because of the varying conditions used to diagnose and evaluate organic or functional anomalies. In addition, it is expected that the fixation stability determined by Scanpaths (capitalized for clarity as the name of a method) related to the gaze path can be evaluated by areas. The Scanpath method is shown as the sequence of positions and connecting lines for gaze points and evaluated as the area surrounded by them.

The purpose of our study was to quantify fixation stability according to image-based areas using the BCEA, KDE, and Scanpath methods, with the help of an eye tracker, in patients with abnormal phoria and to compare the characteristics among methods.

Materials and methods

Participants

This was a prospective cross-sectional study. The minimum required sample size was estimated using G*Power version 3.1 software (Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany). The required sample size was calculated using a power analysis by applying the criteria for classifying normal (mean $= 3\,\Delta$, prism diopter; range $= 0$ to $6\,\Delta$ exophoria) and abnormal phoria (more than $7\,\Delta$ exophoria) and the SD ($\pm 5\,\Delta$) at near. The effect size, α error, power ($1-$ beta), and allocation ratio were 0.8, 0.05, 0.80, and 2, respectively. The required sample size was 19 and 39 participants with normal and abnormal phoria, respectively, for an independent samples t-test. However, in this study, 65 participants (29 women and 36 men; 55% male) between the ages of 19 and 26 years (mean 21.5 ± 1.9 years) were selectively recruited among college students with visual complaints and symptoms and those who arrived for a periodic eye examination at

the Department of Optometry, Kangwon National University.

Twenty participants (9 women and 11 men) and 45 participants (20 women and 25 men) had normal and abnormal phoria, respectively. All participant details were de-identified. The inclusion criteria were normal and abnormal phoria where binocular viewing was possible and the absence of strabismus and amblyopia. Participants with a history of prior surgery or strabismus were also included if they exhibited temporary binocular vision, such as an intermittent or alternating tropia (if not constant strabismus). The included participants had habitual visual acuity ranging from 0.15 logMAR (0.7 decimal) to −0.08 logMAR (1.2 decimal). Patients who exhibited strabismus and in whom binocular viewing was not possible were excluded. Written consent was obtained after a verbal explanation of the nature of the research. This study was approved by the Institutional Review Board of Kangwon National University, Chuncheon, Korea (Approval No. KWNUIRB-2018-10-002-002; date of approval, 21 Jan 2020) and was performed in accordance with the tenets of the Declaration of Helsinki. This prospective study complies with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement guidelines.[19]

Participants with good ocular health and those not taking ocular medications underwent the following preliminary examinations: evaluation of habitual corrected or uncorrected visual acuity, refractive power for the ophthalmic lenses of habitual spectacles, and inter-pupillary distance. The test for phoria was performed using prism bars at near (40 cm) and distance (6 m). Negative and positive values for phoria indicated exophoria and esophoria, respectively. The normal criteria for distance and near phoria are 1 exophoria $\pm 1 \Delta$ and 3 exophoria ± 3 Δ, respectively.[20] The calculated accommodative convergence to accommodation (AC/A) ratio was the sum of the inter-pupillary distance (cm) and the difference between the near and distance phoria divided by 2.50 diopters (D).

All participants were classified into abnormal and normal phoria groups on the basis of their phoria at near and distance and the AC/A ratio according to Scheiman and Wick's study.[21]

Measurement of fixation stability using an eye tracker

According to a procedure described in our previous study,[22] eye movements were recorded using the Clinical Eye Tracker system (Version 18.04, Thomson Software Solutions, Hatfield, UK) equipped with a non-invasive and measurable remote eye tracker. This eye tracking system recorded horizontal (x-axis) and vertical (y-axis) gaze positions with a frequency of 70 Hz. The distance between the display (27 inches, 1920×1080 pixels, $100 \, cd/m^2$) and the eyes was 550 mm. The participants were instructed to place their chin in the chin rest and forehead against the forehead rest to avoid head movements during fixation stability evaluation. The central fixation target was a red dot 3.7 mm in diameter (12 pixels, 0.38 degrees (deg)) corresponding to a visual acuity of 0.66 logMAR (0.22 decimal).

The gaze positions for a participant were selected at random and measured for 15 s during binocular viewing of the target under normal room illumination. The measurement value after two simulations for each participant was recorded without using the average value of repeated measurements. The binocular gaze position was calculated from the mean gaze position of the right and left eyes. Recorded data were exported in an Excel file, and data for analysis were collected for 10 s, from 1 s to 11 s after the start. Blinks were detected by the

Clinical Eye Tracker, and any data with blinks were excluded from the analysis. The collected data were transformed into a Cartesian coordinate (x, y) system so that the center of the display corresponded to the zero point. In this study, pixel units were converted to deg units using a conversion factor of 0.032 deg/pixel as needed.[23]

Experimental protocol

The differences observed in the fixation stability between the abnormal and normal phoria groups were measured for each method. In the absence of differences, the fixation stability for each method in the abnormal phoria group, which had more clinical signs and symptoms, was compared with that in the normal phoria group. As shown in Figure 1, the experimental procedures for fixation stability comparison between the abnormal and normal phoria (control) group and for comparisons among methods using image-based areas were as follows: calculation of the BCEA using the formula (calculated BCEA, CalBCEA), plotting of images obtained by each method using RStudio software
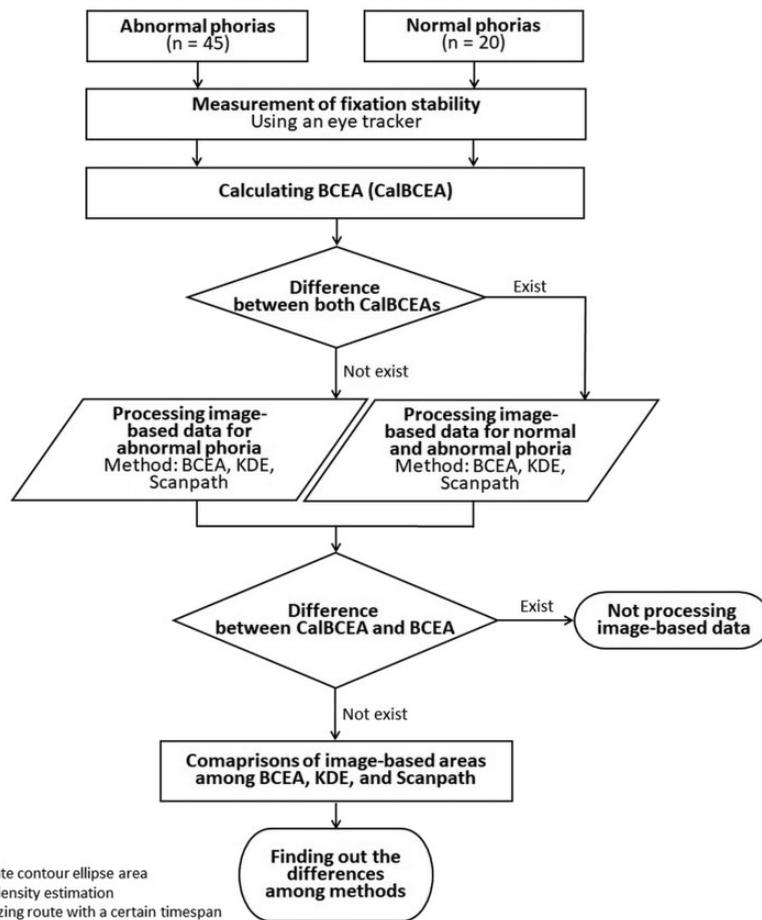
Figure 1. Flow chart for the comparison of fixation stability by image-based areas.

(version 1.3.1093; RStudio, Boston, MA, USA),[24] and measuring the image-based areas using ImageJ software (version 1.52a; NIH, Bethesda, MD, USA).[25] Thus, the fixation stability was evaluated not only by the CalBCEA but also by the BCEA, KDE, and Scanpath methods using image-based areas.

The BCEA encompasses a given proportion of all fixation points and is calculated using a formula relating the SD (σ_h and σ_v), correlation (ρ), and probability area (k) over the horizontal and vertical positions of given points as follows (Eq. 1):

$$BCEA = 2k\pi\sigma_h\sigma_v(1 - \rho^2)^{1/2} \qquad (1)$$

Mathematically, the basic concept for KDE is derived from a univariate distribution with a density function at any given point $(x_1, x_2, \ldots, x_n)$. It consists of a kernel function (K) as the probability density and a bandwidth (h) as a smoothing parameter as follows (Eq. 2):

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \qquad (2)$$

Using the CalBCEA, the fixation stability was calculated using Eq. 1.[12] In this equation, σ_h and σ_v are the SDs of the gaze points in the horizontal (x-axis) and vertical positions (y-axis), respectively; ρ is the product-moment correlation of the two positional components; and k is dependent upon the probability area chosen (P) (Eq. 3).

$$P = 1 - e^{-k} \qquad (3)$$

In Eq. 3, e is the base of the natural logarithm. Therefore, k is 1.146 for a probability of 68.2%, corresponding to ± 1 SD, and is 3.079 for a probability of 95.4%, corresponding to ± 2 SD. A smaller BCEA indicates better fixation stability. Log transformed data were used to perform parametric statistical analyses. The BCEA obtained with the image-based area was compared with the CalBCEA to evaluate their agreement.

The fixation stability of the image-based areas determined by the BCEA, KDE, and Scanpath methods (Figure 2) was plotted using RStudio software and evaluated using ImageJ software. The calculation of areas by ImageJ software was performed as follows: (1) images obtained by each method were plotted without gaze points using RStudio software and transferred to ImageJ software; (2) images were traced using a tracing tool and measured on horizontal and vertical scales (pixels); (3) a pixel aspect ratio and a scale were set for the images, and the area was analyzed. Areas with pixels were converted to areas with deg units using a conversion factor of 0.032 deg/pixel.

KDE is a method used to visualize the probability density function for gaze points of fixation,[15] while the Scanpath, as a new method applied in this study, is an area enclosed by the gaze path for a particular timespan.

Statistical analysis

There were no missing data for any participants. All data were collected and statistically analyzed using MedCalc software (Version 12.7.7.0; MedCalc Software, Mariakerke, Belgium). The Kolmogorov–Smirnov test was used to assess normality. An independent samples t-test was used to compare the mean values of the abnormal and normal phoria groups. A paired t-test was performed for comparisons between two methods in the same group. Repeated-measures analysis of variance (ANOVA) followed by Bonferroni's post-hoc test was performed for
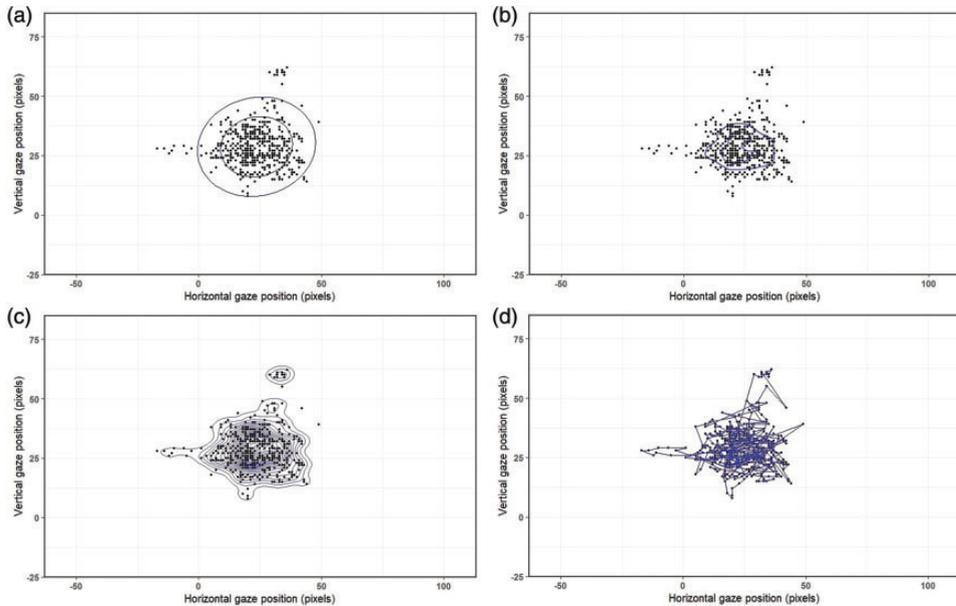
Figure 2. Examples of several fixation stabilities according to image-based area. In this study, the factor used to convert from pixels to angles was 0.032. (a) Bivariate contour ellipse area (BCEA) with 68.2% for the inner ellipse and 95.4% for the outer ellipse. (b) Kernel density estimation (KDE) for 68.2%. (c) KDE for 95.4% and (d) Scanpath. Each fixation stability was evaluated as an area using ImageJ software.

comparisons among methods used to determine image-based areas in abnormal phoria. The reliability and agreement between methods were evaluated using a Bland–Altman analysis. Pearson's correlation coefficient and regression lines were used to evaluate the degree and strength of associations between methods used to determine image-based areas. A p-value less than 0.05 was considered statistically significant.

Results

Differences in demographic and clinical characteristics

Regarding the demographic and clinical characteristics of participants as shown in Table 1, no significant differences in mean values were observed between the abnormal (n = 45) and normal phoria (n = 20) groups,

except in distance phoria, near phoria, and ages. Furthermore, unlike the normal phoria group, the abnormal phoria group had symptoms and signs of near binocular anomalies, including convergence insufficiency (n = 24), convergence excess (n = 9), basic exophoria, and esophoria (n = 8 and 4, respectively).

Comparisons between BCEAs in phoria groups

Evaluation of the fixation stability using an image-based area can be applied to the forms plotted by various measurement methods, especially Scanpaths, which cannot be performed by calculation methods. To verify the possibility of the application of image-based areas for evaluation of fixation stability, we investigated whether any difference existed between the CalBCEA and the BCEA using

Table 1. Demographic and clinical characteristics of participants.

Parameter	Abnormal phoria (n = 45)	Normal phoria (n = 20)	p-value
Sex (male/female, n)	25/20	11/9	
Age (years)	21.93 ± 1.98	20.65 ± 1.27	$t_{63} = -2.658$, 0.010
Visual acuity (logMAR)			
Right eye	0.02 ± 0.04	0.03 ± 0.06	$t_{63} = 1.250$, 0.216
Left eye	0.01 ± 0.04	0.02 ± 0.05	$t_{63} = 0.236$, 0.814
Optical correction (diopters)			
Spherical equivalent for right eye	−4.54 ± 2.55	−4.39 ± 2.97	$t_{48} = 0.174$, 0.863
Spherical equivalent for left eye	−4.16 ± 2.78	−4.55 ± 2.11	$t_{47} = -0.453$, 0.653
Phoria (prism diopters)			
Distance	−3.44 ± 5.00	−0.85 ± 0.81	$t_{63} = 2.296$, 0.025
Near	−8.73 ± 5.53	−3.60 ± 2.33	$t_{63} = 2.368$, 0.021
Calculated AC/A	4.26 ± 2.19	5.24 ± 0.99	$t_{63} = 1.911$, 0.061

Data are presented as the mean ± standard deviation and the number of participants.
Minus and plus signs for phoria indicate exophoria and esophoria, respectively.
Calculated accommodative convergence to accommodation (calculated AC/A).
The p-values were determined using an independent samples t-test.

Table 2. Fixation stability based on the calculated bivariate contour ellipse area in the two phoria groups.

	Abnormal phoria (n = 45)	Normal phoria (n = 20)
Mean ± SD (95% CI), deg^2		
CalBCEA1SD	0.69 ± 0.56 (0.52–0.86)	0.67 ± 0.38 (0.39–1.06)
CalBCEA2SD	1.86 ± 1.50 (1.41–2.31)	1.80 ± 1.02 (1.05–2.83)
Mean ± SD (95% CI), log deg^2		
CalBCEA1SD	−0.28 ± 0.31 (−0.37 to −0.18)	−0.25 ± 0.27 (−0.41 to 0.02)
CalBCEA2SD	0.15 ± 0.31 (0.06–0.25)	0.18 ± 0.27 (0.02–0.45)
Independent samples t-test	CalBCEA1SD between groups: t = −0.35 (p = 0.729)	
	CalBCEA2SD between groups: t = −0.34 (p = 0.738)	

CalBCEA1SD and CalBCEA2SD = calculated BCEA for ±1 SD (68.2%) and ±2 SD (95.4%), SD = standard deviation, CI = confidence interval, BCEA = bivariate contour ellipse area.

image-based areas from the two groups. First, we compared the fixation stability between the two groups and then evaluated the interchangeability of the CalBCEA and BCEA.

The means and deviations of CalBCEAs for the abnormal and normal phoria groups are shown in Table 2. The mean values of the abnormal and normal groups were −0.28 log deg^2 (0.69 deg^2) and −0.25 log deg^2 (0.67 deg^2) for ±1 SD (68.2%) as expressed by the CalBCEA1SD and 0.15 log deg^2 (1.86 deg^2) and 0.18 log deg^2 (1.80 deg^2) for ±2 SD (95.4%) as indicated by the CalBCEA2SD. There were no significant differences in the CalBCEA1SD or CalBCEA2SD between the two groups.

Comparison of the CalBCEA and BCEA in the abnormal phoria group, which showed a larger degree of phoria than the control group, yielded the following results. A paired t-test showed no significant difference between CalBCEA1SD (−0.28 ± 0.31 log deg^2) and BCEA1SD (−0.28 ± 0.31 log

deg^2) and between the CalBCE2SD (0.15 ± 0.31 log deg^2) and BCEA2SD (0.15 ± 0.32 log deg^2) (t = 1.50, p = 0.141 for ±1 SD; t = 0.58, p = 0.568 for ±2 SD). Pearson's correlation coefficient between the CalBCEA1SD and BCEA1SD and that between the CalBCEA2SD and BCEA2SD indicated strong positive correlations of 0.999 (p < 0.0001) and 0.997 (p < 0.0001), respectively.

The agreement between the CalBCEA and BCEA is shown in Figure 3 as Bland–Altman plots. The CalBCEA1SD vs BCEA1SD and CalBCEA2SD vs BCEA2SD showed mean differences of 0.003 and 0.002 log deg^2, respectively, indicating a very small systematic error, and the 95% limits of agreement were −0.02 to 0.03 and −0.05 to 0.05, respectively, indicating a small range of error that may be clinically acceptable. Hence, the CalBCEA and BCEA with 1 SD and 2 SD highly agree and can be used interchangeably.

Comparisons between methods in the abnormal phoria group

Comparisons between methods to determine fixation stability using image-based areas in the abnormal phoria group are shown in Table 3.

The comparative results of methods to determine image-based areas in the abnormal phoria group are shown in Table 4. The means (95% CI for mean) after log$_{10}$ transformation were −0.28 log deg^2 (−0.37 to −0.18) for BCEA1SD, −0.54 log deg^2 (−0.63 to −0.45) for KDE1SD, 0.15 log deg^2 (0.06–0.25) for BCEA2SD, 0.04 log deg^2 (−0.06 to 0.13) for KDE2SD, and 0.05 log deg^2 (−0.05 to 0.16) for Scanpath. Repeated measures ANOVA indicated that the differences among the image-based areas (log deg^2) were significant (F = 228.68, p < 0.001). Post-hoc tests with Bonferroni corrections indicated that differences of −0.26 log deg^2 between KDE1SD and BCEA1SD, −0.12 log deg^2 between KDE2SD and BCEA2SD, and −0.10 log deg^2 between Scanpath and BCEA2SD were significant (p < 0.001 for all). The difference of −0.02 log deg^2 between Scanpath and KDE2SD was not significant. According to the results, the image-based areas, indicating the size of fixation, increased in the following order: KDE1SD, BCEA1SD, KDE2SD or Scanpath, and BCEA2SD.
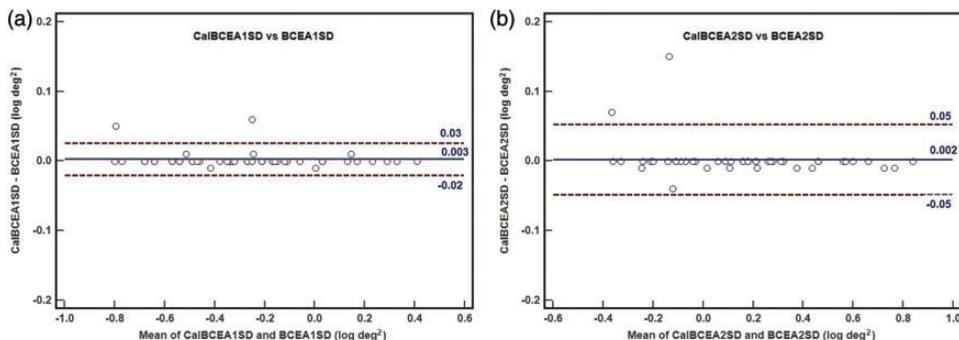
Figure 3. Comparisons of the calculated bivariate contour ellipse area (CalBCEA) and bivariate contour ellipse area (BCEA) in participants with abnormal phoria. The dotted lines represent the mean CalBCEA and BCEA with the upper and lower 95% limits of agreement (mean difference ±1.96 × standard deviation (SD) of the difference). (a) CalBCEA1SD and BCEA1SD ±1 SD and (b) CalBCEA2SD and BCEA2SD ±2 SD.

Table 3. Comparisons among the different methods.

Difference (deg^2)	Positive value (n)	Negative value (n)	Equal value (n)
BCEA1SD – KDE1SD	41	3	1
BCEA2SD – KDE2SD	41	4	
BCEA2SD – Scanpath	35	9	1
KDE2SD – Scanpath	22	23	

BCEA1SD and BCEA2SD = image-based bivariate contour ellipse area for ±1 standard deviation (SD) and ±2 SD, KDE1SD and KDE2SD = image-based area determined by kernel density estimation for ±1 SD and ±2 SD.

Table 4. Comparisons among various methods to determine image-based areas.

	Mean ± SD	(95% CI), deg^2	Mean ± SD	z (95% CI), log deg^2	F, p-value (*post-hoc*)
a. BCEA1SD	0.69 ± 0.56	(0.52–0.86)	−0.28 ± 0.31	(−0.37 to −0.18)	F = 228.68, p < 0.001
b. KDE1SD	0.36 ± 0.25	(0.28–0.43)	−0.54 ± 0.29	(−0.63 to −0.45)	(c > d, e > a > b)
c. BCEA2SD	1.86 ± 1.51	(1.40–2.31)	0.15 ± 0.32	(0.06–0.25)	
d. KDE2SD	1.44 ± 1.29	(1.05–1.83)	0.04 ± 0.32	(−0.06 to 0.13)	
e. Scanpath	1.54 ± 1.31	(1.15–1.93)	0.05 ± 0.35	(−0.05 to 0.16)	

BCEA1SD and BCEA2SD = image-based BCEA for ±1 SD and ±2 SD, KDE1SD and KDE2SD = image-based area by the kernel density estimation for ±1 SD and ±2 SD, SD = standard deviation, CI = confidence interval, BCEA = bivariate contour ellipse area , KDE = kernel density estimation.
The p-values were determined using repeated measures analysis of variance followed by Bonferroni's post-hoc test.

Relationship and agreement among methods

In comparisons between different methods for determining the image-based area (log deg^2) in the abnormal phoria group, the Pearson's correlation coefficients exhibited high positive correlations between KDE1SD and BCEA1SD (0.71, p < 0.0001), KDE2SD and BCEA2SD (0.89, p < 0.0001), Scanpath and BCEA2SD (0.94, p < 0.0001), and Scanpath and KDE2SD (0.81, p < 0.0001).

In scatter diagrams with regression lines (Figure 4), the regression line obtained using ANOVA was statistically significant in comparisons of KDE1SD and BCEA1SD (F = 44.48, p < 0.001), KDE2SD and BCEA2SD (F = 172.94, p < 0.001), Scanpath and BCEA2SD (F = 301.15, p < 0.001), and Scanpath and KDE2SD (F = 82.74, p < 0.001). The coefficients of determination between KDE1SD and BCEA1SD, KDE2SD and BCEA2SD, Scanpath and BCEA2SD, and Scanpath and KDE2SD were 0.509, 0.801, 0.875, and 0.657, respectively. KDE2SD and Scanpath accounted for 80.1% and 87.4% of the BCEA2SD value, respectively. However, 50.8% of the BCEA1SD value was determined by KDE1SD, and 65.6% of the KDE2SD value was determined by Scanpath.

In the Bland–Altman plots for agreement analysis between different methods of determining the image-based areas in the abnormal phoria group (Figure 5), the bias was lowest for the mean difference between Scanpath and KDE2SD (0.02 log deg^2) and highest for the mean difference between KDE1SD and BCEA1SD (−0.26 log deg^2). The bias between KDE1SD and BCEA1SD was not significant because the line of equality (zero) was within the interval of the mean difference. The 95% limits
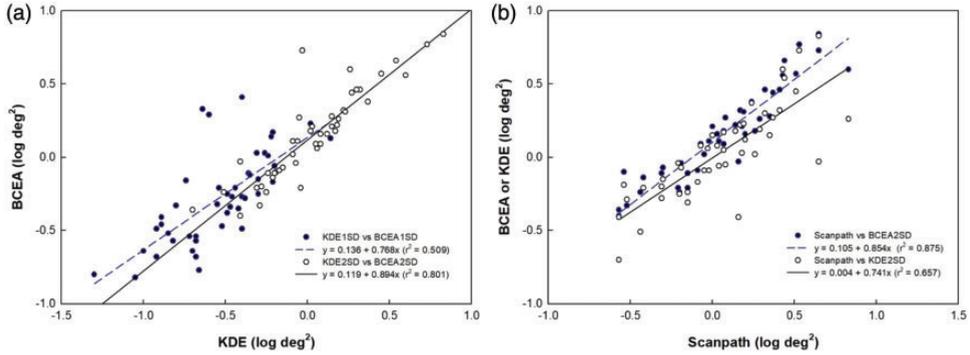
Figure 4. Regression lines among various methods of determining image-based areas. (a) Kernel density estimation (KDE) with 1 standard deviation (SD) vs. bivariate contour ellipse area (BCEA) with 1 SD (filled circles), and KDE2SD vs. BCEA2SD (open circles) and (b) Scanpath vs. BCEA2SD (filled circles) and Scanpath vs. KDE2SD (open circles).
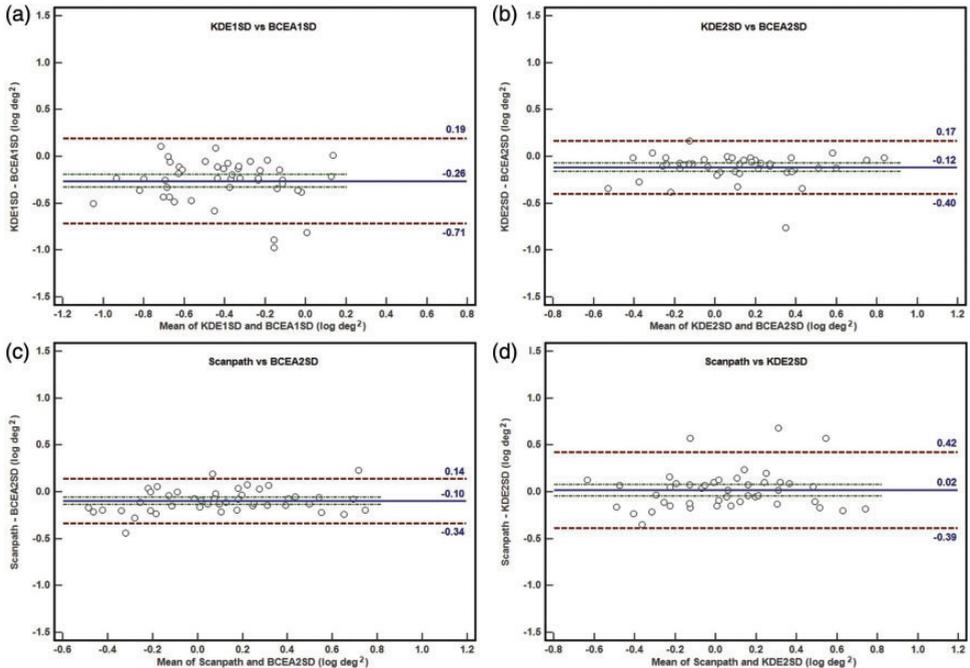
Figure 5. Bland–Altman plots comparing methods for determining the image-based areas. (a) Kernel density estimation with 1 standard deviation (KDE1SD) and bivariate contour ellipse area with 1 standard deviation (BCEA1SD). (b) KDE2SD and BCEA2SD. (c) Scanpath and BCEA2SD and (d) Scanpath and KDE2SD. The dotted lines represent the mean calculated BCEA and BCEA with the upper and lower 95% limits of agreement (mean difference $\pm 1.96 \times$ SD of the difference). Green dotted/broken lines indicate the 95% confidence interval for mean differences.

of agreement between methods increased in the following order: Scanpath and BCEA2SD (-0.10 ± 0.24 log deg^2), KDE2SD and BCEA2SD (-0.12 ± 0.28 log deg^2), Scanpath and KDE2SD (0.02 ± 0.40 log deg^2), and KDE1SD and BCEA1SD (0.26 ± 0.45 log deg^2).

Discussion

The difference between the BCEA and KDE is that in the former, only the probability based on the overall shape is considered, while the latter is determined by the probability density. In contrast, Scanpath is a method based on traces of points over time and not on a mathematical formula. However, because of differences in conditions, relative comparisons between these methods are not easy. No study to date has characterized the ability of image-based quantification to evaluate fixation stability in abnormal phoria with binocular vision without strabismus and amblyopia.

Specifically, if each method is characterized on the basis of SDs, density functions, and traces of points over time, then fixation stability determined using image-based areas in abnormal phoria will show similar characteristics and significant differences among the BCEA, KDE, and Scanpath methods of evaluation. Phoria for non-strabismic binocular anomalies cannot be detected by clinical tests such as a direct observation during binocular viewing. Eye movements using eye-tracking systems could be a new clinical test for evaluation of phorias. Fixation stability may vary with characteristics of the presenting anomaly. Fixation on gaze distance may be centralized or delocalized depending on phoria at distance and near, fusional vergence, and the AC/A ratio for abnormal phorias such as convergence excess, convergence insufficiency, divergence excess, divergence insufficiency, basic exophoria, basic esophoria, and other binocular anomalies. In this study, however, image-based evaluations of fixation stability were compared as a whole and were not based on each characteristic as a means to classify binocular vision disorders for individuals.

Reliability of image-based quantification

The BCEA, quantified as an image-based area, can be interchanged with the CalBCEA. Although a mean difference in phoria existed between the two groups, the mean difference in fixation stability, as determined by the CalBCEA, was not significantly different between the two groups.

Fixation stability varied according to the measurement conditions in several studies. For eye position stability in amblyopia and in normal binocular vision,[26] the binocular fixation stability (-0.88 log deg^2) for normal binocular vision was better than that for monocular fixation stability (-0.59 log deg^2), while that for the amblyopic eye was -0.44 log deg^2. Another study[27] reported that while the best fixation stability was evident in healthy controls (average -0.46 log deg^2 for monocular viewing conditions), the fixation instability increased with increasing severity of amblyopia. In a previous study of monocular viewing,[28] fixation stability with ± 1 SD in strabismus without binocular vision was larger for both the non-preferred and preferred eyes (0.10 log deg^2 and -0.17 log deg^2, respectively) than that in our study. Many previous studies showed that fixation under binocular viewing conditions is more stable than that in monocular viewing conditions.[26,29,30]

With regard to the BCEA values for binocular viewing, our results exhibited decreased fixation stability compared with those of previous studies. This lower fixation stability could not be explained clearly except by the differences in observer and experimental conditions, but a relative comparison among the methods for evaluation

of fixation stability of binocular vision was possible.

The correlation between the CalBCEA and BCEA values for fixation stability was evaluated in the abnormal phoria group. Both the Pearson's correlation coefficient, as a measure of relative reliability, and Bland-Altman analysis, as a measure of absolute reliability,[31] revealed that the BCEA and CalBCEA can be used interchangeably for the quantification of fixation stability.

Comparison of characteristics among image-based methods

In the comparison of the characteristics of the three image-based area methods for evaluating fixation stability in the abnormal phoria group, the BCEA was relatively overestimated because of the inclusion of the space without gaze positions, the KDE was relatively underestimated because of the influence of gaze density, and Scanpath showed a tendency to be close to the KDE2SD value (KDE with ± 2 SD).

A previous study on fixation stability[32] reported that it is difficult to compare fixation stability across studies because of the different methodologies used for data collection and quantification. Therefore, the study suggested that a single metric can be used to unify the quantification of fixation stability for data obtained using different eye-trackers, facilitating the comparison of fixation stability across studies. Another evaluation of fixation stability based on the BCEA[33] reported that the calibration method, operator, participant's eye physiology, and visual aids affect the quality of data recorded with a video-based eye tracker. Unlike the aforementioned studies, our study quantified and compared image-based methods including the BCEA, KDE, and Scanpath. In our study, it is reasonable that the mean differences between methods, such as those with ± 2 SD being

greater than those with ± 1 SD,[12] indicated that the BCEA2SD value was greater than the BCEA1SD value and the KDE2SD value was greater than the KDE1SD value.

The Scanpath value was greater than the BCEA1SD and KDE1SD values and was between the KDE2SD and BCEA2SD values, suggesting that areas determined by traces of points over time are usually determined between areas based on SDs and density functions.[12,15] In the evaluation of the degree of association between methods, the Scanpath method showed a very high positive relationship with the BCEA2SD. Therefore, regarding the coefficient of determination denoted by r^2, the BCEA2SD result was highly explained because Scanpath and KDE1SD (2 SD) had a high positive correlation with BCEA1SD (2 SD). These results suggest that some similarity exists between the two methods in determining the probability area and the probability density.[12,15] A high positive correlation and a good fit of linear regression in Scanpath vs BCEA2SD do not signify the extent of agreement but provide information about the degree of association between them.[34]

In Bland–Altman plots, Scanpath was similar to KDE2SD for systematic error, representing bias, but also similar to BCEA2SD in 95% limits of agreement, which is related to clinical acceptability. KDE1SD was highly different from BCEA1SD in both systematic error and the 95% limits of agreement. The clinically acceptable value for fixation stability is unknown, but the value is assumed to be ± 0.40 log deg^2 ($1.96 \times$ SD) for the 95% limits of agreement when calculated based on the SD of mean values of -0.88 log deg^{2,26} -0.35 log deg^{2,28} -0.39 log deg^{2,28} -0.24 log deg^{2,34} -0.77 log deg^{2,18} -0.52 log deg^{2,18} -0.68 log deg^{2,13} -0.52 log deg^{2,35} and -0.57 log deg^{2,36} for control or normal groups in previous studies with different conditions such as binocular

anomalies,[21] longer fixations,[27] and various stimuli.[36] In analyses based on clinical acceptance, it is not easy to choose one among these three methods, even if bias is taken into account.

In a study evaluating fixational eye movements in people with macular disease,[37] the BCEA and the KDE, as isolines that correspond to 68% in the probability density function, agree reasonably well in representing fixation stability. However, our findings showed that at the same probability (±1 SD, ±2 SD), the BCEA tended to be larger than the KDE, and at higher probability, the BCEA was also larger. The Scanpath was affected by the sequence of gaze paths, and the mean differences were less with KDE2D than with BCEA2SD, but the 95% limits of agreement were larger with KDE2D than with BCEA2SD. In cases of crowded gaze points, the KDE was small; therefore, the fixation stability was increased. In cases with dispersed gaze points, the KDE appeared as isolated contours.[15] A larger Scanpath than BCEA2SD indicated that the path of gaze departed largely and frequently from crowded gaze positions. Furthermore, regarding image-based methods, the BCEA tended to show a wide area because a data-free space was included, which can simplify the main direction of gaze. The KDE appeared as a narrower area for ±1 SD and occasionally as isolated contours for ±2 SD because of the density of gaze points, but this method cannot simplify the main direction of gaze similar to the BCEA. Unlike the BCEA and KDE, which show patterns regardless of time, the Scanpath represents an area that spans all gaze points over time and can be applied to ±2 SD but not to ±1 SD, i.e., it cannot be applied with various probabilities. Unlike other methods, it is difficult to find the density of gaze points and the main direction of gaze with the Scanpath.

This study had some limitations. First, our study was limited to patients with abnormal phoria with binocular vision. Therefore, fixation stability was not poor, and distinct differences distinguishing it from normal phoria, such as strabismus without binocular vision,[38] were not observed. Therefore, further studies are needed to quantify fixation stability using image-based areas in patients with amblyopia and strabismus, as well as in patients with various binocular vision anomalies. Second, fixation stability was evaluated over a short interval; therefore, the variability depending on time, such as more than 30 s,[13,39] was not evaluated. Finally, although our study was conducted on participants without eye diseases, mental illness-related fixation stability is a variable that remains to be evaluated in the future, as phoria is affected by the convergence deficits in schizophrenia,[40] and eye movement deficits are present among patients with schizophrenia.[41] Regardless of these limitations, this study showed various applications of the evaluation of fixation stability using image-based areas.

In summary, the clinical introduction of eye trackers in ophthalmology and optometry has enabled improved evaluation of eye movements and visual performance to develop a means to prevent, diagnose, and treat abnormalities or ocular disease, such as cataract, glaucoma, amblyopia, and strabismus. In this study, we examined and compared the use of image-based quantification via several methods of evaluating fixation stability using an eye tracker. This image-based quantification can be applied to various types of fixation analyses such as the BCEA, KDE, and Scanpath methods. The BCEA method of quantifying fixation stability was relatively overestimated because of the inclusion of the space without gaze positions; the KDE was relatively underestimated because of the influence of gaze density; and Scanpath showed a tendency to be similar to KDE2SD (KDE with ±2 SD). Fixation stability should be

evaluated while considering the characteristics of each method, and it is recommended to use more than one method.

Author contributions

Contributions to this study: concepts and design (DSY; SYK), literature search (DSY; SYK; HGC), clinical and experimental studies (DSY; SYK), data acquisition and analysis (DSY; SYK; BYM; HGC), manuscript preparation and editing (DSY; SYK; BYM), manuscript review (BYM; HGC). All authors approved the final manuscript.

Declaration of conflicting interest

The authors declare that there is no conflict of interest.

Funding

ORCID iD

Dr Dong-Sik Yu https://orcid.org/0000-0002-4387-4408

References

1. Mihara M, Hayashi A, Fujita K, et al. Fixation stability of the upward gaze in patients with myasthenia gravis: an eye-tracker study. *BMJ Open Ophthalmol* 2017; 2: e000072.

2. Molina-Martín A, Piñero DP and Pérez-Cambrodí RJ. Fixation pattern analysis with microperimetry in nystagmus patients. *Can J Ophthalmol* 2015; 50: 413–421.

3. Subramanian V, Jost RM and Birch EE. A quantitative study of fixation stability in amblyopia. *Invest Ophthalmo Vis Sci* 2013; 54: 1998–2003.

4. Tarita-Nistor L, González EG, Brin T, et al. Fixation stability and viewing distance in patients with AMD. *Optom Vis Sci* 2017; 94: 239–245.

5. Wan Y, Yang J, Ren X, et al. Evaluation of eye movements and visual performance in patients with cataract. *Sci Rep* 2020; 10: 9875.

6. Ciuffreda MA, Ciuffreda KJ and Santos D. Visagraph: baseline analysis and procedural guidelines. *J Behav Optom* 2003; 14: 60–64.

7. Reddy AVC, Mani R, Selvakumar A, et al. Reading eye movements in traumatic brain injury. *J Optom* 2020; 13: 155–162.

8. Cognolato M, Atzori M and Müller H. Head-mounted eye gaze tracking devices: an overview of modern devices and recent advances. *J Rehabil Assistive Technol Eng* 2018; 5: 1–13.

9. Kasneci E, Black AA and Wood JM. Eye-tracking as a tool to evaluate functional ability in everyday tasks in glaucoma. *J Ophthalmol* 2017; 2017: 6425913.

10. Di Russo F, Pitzalis S and Spinelli D. Fixation stability and saccadic latency in élite shooters. *Vision Res* 2003; 43: 1837–1845.

11. Rao RP, Zelinsky GJ, Hayhoe MM, et al. Eye movements in iconic visual search. *Vision Res* 2002; 42: 1447–1463.

12. Crossland MD and Rubin GS. The use of an infrared eyetracker to measure fixation stability. *Optom Vis Sci* 2002; 79: 735–739.

13. Fragiotta S, Carnevale C, Cutini A, et al. Factors influencing fixation stability area: a comparison of two methods of recording. *Optom Vis Sci* 2018; 95: 384–390.

14. Castet E and Crossland M. Quantifying eye stability during a fixation task: a review of definitions and methods. *Seeing Perceiving* 2012; 25: 449–469.

15. Crossland MD, Sims M, Galbraith RF, et al. Evaluation of a new quantitative technique to assess the number and extent of preferred retinal loci in macular disease. *Vision Res* 2004; 44: 1537–1546.

16. Crossland MD, Dunbar HM and Rubin GS. Fixation stability measurement using the MP1 microperimeter. *Retina* 2009; 29: 651–656.

17. Montesano G, Crabb DP, Jones PR, et al. Evidence for alterations in fixational eye movements in glaucoma. *BMC Ophthalmol* 2018; 18: 191.

18. Ghasia FF, Otero-Millan J and Shaikh AG. Abnormal fixational eye movements in strabismus. *Br J Ophthalmol* 2018; 102: 253–259.

19. Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147: 573–577.

20. Moon B-Y, Kim S-Y and Yu D-S. Receiver operating characteristic curve analysis of clinical signs for screening of convergence insufficiency in young adults. *PLoS One* 2020; 15: e0228313.

21. Darko-Takyi C, Khan NE and Nirghin U. A review of the classification of nonstrabismic binocular vision anomalies. *Optometry Reports* 2016; 5: 5626.

22. Kim S-Y, Moon B-Y, Cho HG, Yu D-S. Quantitative evaluation of the association between fixation stability and phoria during short-term binocular viewing. *Front Neurosci* 2022; 16: 721665.

23. Kar A and Corcoran P. Performance evaluation strategies for eye gaze estimation systems with quantitative metrics and visualizations. *Sensors (Basel)* 2018; 18: 3151.

24. RStudio Team. RStudio: Integrated development for R. Boston, MA: RStudio, Inc. 2015. http://www.rstudio.com/ (accessed 20 October 2020).

25. Ferreira T and Rasband W. ImageJ user guide IJ 1.45r, 2012. https://imagej.nih.gov/ij/docs/guide/user-guide.pdf (accessed 12 March 2021).

26. González EG, Wong AM, Niechwiej-Szwedo E, et al. Eye position stability in amblyopia and in normal binocular vision. *Invest Optom Vis Sci* 2012; 53: 5386–5394.

27. Shaikh AG, Otero-Millan J, Kumar P, et al. Abnormal fixational eye movements in amblyopia. *PLoS One* 2016; 11: e0149953.

28. Kelly KR, Cheng-Patel C, Jost RM, et al. Fixation instability during binocular viewing in anisometropic and strabismic children. *Exp Eye Res* 2019; 183: 29–37.

29. Raveendran RN, Bobier WR and Thompson B. Binocular vision and fixational eye movements. *J Vis* 2019; 19: 9.

30. Motter BC and Poggio GF. Binocular fixation in the rhesus monkey: spatial and temporal characteristics. *Exp Brain Res* 1984; 54: 304–314.

31. Bruton A, Conway JH and Holgate ST. Reliability: What is it and how is it measured? *Physiotherapy* 2000; 86: 94–99.

32. Giavarina D. Understanding Bland Altman analysis. *Biochem Med* 2015; 25: 141–151.

33. Chung S, Ağaoğlu M, Krishnan A. Unifying the quantification of fixation stability. In: *Journal of Vision* Abstract from Vision Sciences Society Eighteenth Annual Meeting St. Pete Beach, FL, USA, 18–23 May 2018; 18: 1000.

34. Nyström M, Andersson R, Holmqvist K, et al. The influence of calibration method and eye physiology on eyetracking data quality. *Behav Res Methods* 2013; 45: 272–288.

35. Liu H, Bittencourt MG, Sophie R, et al. Fixation stability measurement using two types of microperimetry devices. *Trans Vis Sci Technol* 2015; 4: 3.

36. Hirasawa K, Okano K, Koshiji R, et al. Smaller fixation target size is associated with more stable fixation and less variance in threshold sensitivity. *PLoS One* 2016; 11: e0165046.

37. Kumar G and Chung ST. Characteristics of fixational eye movements in people with macular disease. *Invest Ophthalmol Vis Sci* 2014; 55: 5125–5133.

38. Bui Quoc E and Milleret C. Origins of strabismus and loss of binocular vision. *Front Integr Neurosci* 2014; 8: 71.

39. Raveendran RN, Babu RJ, Hess RF, et al. Transient improvements in fixational stability in strabismic amblyopes following bifoveal fixation and reduced interocular suppression. *Ophthalmic Physiol Opt* 2014; 34: 214–225.

40. Bolding MS, Lahti AC, Gawne TJ, et al. Ocular convergence deficits in schizophrenia. *Front Psychiatry* 2012; 3: 86.

41. Wolf A, Ueda K and Hirano Y. Recent updates of eye movement abnormalities in patients with schizophrenia: A scoping review. *Psychiatry Clin Neurosci* 2021; 75: 82–100.