

1        **Automatic Measurements of Smooth Pursuit Eye Movements by**  
2        **Video-Oculography and Deep Learning-Based Object Detection**

3

4        **Authors:** Masakazu Hirota<sup>1,2</sup>, Takao Hayashi<sup>1,2</sup>, Emiko Watanabe<sup>2</sup>, Atsushi Mizota<sup>2</sup>

5

6        **Affiliations:**

7        <sup>1</sup>Department of Orthoptics, Faculty of Medical Technology, Teikyo University, Itabashi,  
8        Tokyo, Japan.

9        <sup>2</sup>Department of Ophthalmology, School of Medicine, Teikyo University, Itabashi, Tokyo,  
10        Japan.

11

12        **Corresponding author:** Masakazu Hirota

13        Address: 2-11-1 Kaga, Itabashi, Tokyo, 173-8605 Japan

14        Tel: +81-03-3964-1328

15        Fax: +81-03-3963-0303

16        E-mail: [hirota.ortho@med.teikyo-u.ac.jp](mailto:hirota.ortho@med.teikyo-u.ac.jp)

17

18        **Running title**

19        Video-Oculography and Deep Learning-Based Object Detection

20 **Abstract**

21 **Purpose**

22 The purpose of this study was to develop a technique combining video oculography  
23 (VOG) with single shot multibox detector (SSD) to accurately and quantitatively  
24 examine eye movements.

25

26 **Methods**

27 Eleven healthy volunteers ( $21.3 \pm 0.9$  years) participated in this study. Eye  
28 movements were recorded while tracking a target using a custom-made eye tracker.  
29 The subjects were asked to fixate their focus on the nose of the rabbit-like target  
30 (visual angle was  $0.1^\circ$ ), which was manually moved to a distance of 1 meter by the  
31 examiner during the eye movement test. The test produced 500 images from the  
32 VOG external camera and these images were divided into 3 groups (300, 100, and  
33 100) for training, verification, and testing. The performance of the SSD was evaluated  
34 with 75% average precision ( $AP_{75}$ ), and the relationship between the location of the  
35 fixated target (calculated by the SSD) and the positions of both eyes (recorded by the  
36 VOG) was analyzed.

37

38 **Results**

39 The AP<sub>75</sub> of the SSD on one class of targets was 97.7%. The horizontal and vertical  
40 target locations significantly and positively correlated with the horizontal and vertical  
41 both eye positions (*adjusted R*<sup>2</sup> ≥ 0.955, *P* < 0.001).

42

### 43 **Conclusions**

44 Our findings suggest that VOG with SSD is suitable for the evaluation of eye version  
45 movements in standard clinical assessments.

46

### 47 **Translational Relevance**

48 The combination of VOG and SSD can be used to evaluate the SPEM, and this  
49 method can be translated into clinical settings without changing the testing methods.

50

51

52

53

54

55

56

57

## 58 Introduction

59 Strabismus is an eye movement disorder, and it is important to detect  
60 limitation of eye movement by confirming the version eye movement with smooth  
61 pursuit eye movement (SPEM)<sup>1-8</sup> in nine directions.<sup>9-11</sup> In most ophthalmology clinics,  
62 the examiner visually evaluates SPEM.<sup>12, 13</sup> The examiner detects abnormalities in  
63 eye movements by carefully examining SPEM.<sup>12, 13</sup> However, the current method  
64 does not allow recording them during the examination and does not quantitatively  
65 evaluate the SPEM.

66 In the laboratory, several methods have been used to quantify eye  
67 movements, for example, the search coil method,<sup>14, 15</sup> electrooculography,<sup>16</sup> and  
68 video-oculography (VOG).<sup>15</sup> These techniques can record eye movements of both  
69 eyes simultaneously while the subjects are have fixated their focus on a moving  
70 target. In particular, VOG is used to evaluate eye movements in strabismus patients  
71 with wide range of age groups from children to adults<sup>17-20</sup> because it can measure  
72 eye movements non-invasively and easily. However, laboratory methods have not  
73 been introduced in clinical practice for the following two basic reasons: first, the type  
74 of target is restricted. To achieve accuracy at the laboratory level, it is necessary to  
75 present a predetermined target to the subjects according to the programming codes.  
76 In children, the examiner uses a target, such as an anime or game character, to help

77 keep their attention.<sup>10, 18, 21-23</sup> Second, the VOG is operated with the display, which  
78 limits the flexibility of the examination. The target only makes a fixed movement, and  
79 the movement cannot be changed flexibly depending on the eye movement disorder.

80 We hypothesized that a combination of VOG and a deep learning-based  
81 object detection algorithm is one method of translating laboratory methods to the  
82 clinic. Deep learning-based object detection technology can predict the location and  
83 types of objects in one image.<sup>24-26</sup> Furthermore, the algorithm of deep learning-based  
84 object detection can detect the objects in real-time in the movie with a processing  
85 speed of >30 fps. This processing speed is faster than that of conventional simpler  
86 algorithms that use raster scans per image.<sup>27</sup> Therefore, we expected that by using  
87 VOG to record SPEM and deep learning-based object detection to record movements  
88 of the target, the combination system can measure the SPEM and target  
89 simultaneously without changing the clinical examination.

90 Thus, this study aimed to develop a technique that would combine VOG with  
91 deep learning-based object detection that can quantify SPEM. We evaluated the  
92 target location and both eye positions in healthy volunteers without changing the  
93 clinical method. The experiments were designed to be performed as close to the  
94 ophthalmological clinical setting as possible.

95

## 96 **Methods**

### 97 **Subjects**

98           Eleven volunteers (age,  $21.3 \pm 0.9$  years [mean  $\pm$  standard deviation])  
99 participated in this study. All subjects underwent complete ophthalmologic  
100 examinations, including determination of ocular dominance using the hole-in-card test,  
101 best-corrected visual acuity at distance (5.0 m), near point of convergence,  
102 stereoscopic acuity (Titmus stereotest; Stereo Optical Co., Inc., Chicago, USA) at  
103 40 cm, and heterophoria by alternate cover test both at near (33 cm) and at distance  
104 (5.0 m), and fundus examinations. Stereoacuity was converted to the logarithm of the  
105 arcsecond (log arcsec). Participants were excluded if they had a refractive error  
106  $>\pm 10.0$  D.

107           Informed consent was obtained from all subjects after the nature and possible  
108 complications of the study were explained to them. This investigation adhered to the  
109 tenets of the World Medical Association Declaration of Helsinki. Furthermore,  
110 informed consent for publication of identifying information/images in an online  
111 open-access publication was obtained from a part of subjects. The Institutional  
112 Review Board of Teikyo University approved the experimental protocol and consent  
113 procedures (approval no. 18–161).

114

## 115 **Eye movement recordings**

116 Eye movements were recorded during the tracking of the target using a  
117 custom-made eye tracker based on EMR-9 (NAC Image Technology Inc., Tokyo,  
118 Japan) (Fig. 1). The EMR-9 tracker determined the eye positions by detecting the  
119 corneal reflex and pupil center created by the reflection of a near-infrared light that  
120 can be adjusted by the half-mirror. The sampling rate was 240 Hz, the measurement  
121 angle was  $\pm 31^\circ$  from the center of the scene camera, and the measurement error was  
122  $0.2^\circ$ – $0.5^\circ$  (interquartile range) at a distance of 1.0 m. The EMR-9 controller merged  
123 the gaze of both eyes to the real scenes (resolution,  $640 \times 480$  pixels) that were  
124 recorded by a scene camera with a sampling rate of 29.97 Hz and a delay of  $\leq 52$  ms.

125 All subjects underwent a calibration test under binocular conditions with the  
126 fully corrected glasses at 1.0 m before the eye movement test. Noise caused by lens  
127 reflection was noted in 2 of 11 participants and avoided by manually changing the tilt  
128 of the half-mirror. During the calibration, all subjects were asked to fixate a black  
129 nine-point target (visual angle,  $0.1^\circ$ ) on a whiteboard. The center of the screen was  
130 defined as  $0^\circ$ , the right and upper halves of the screen were defined as the positive  
131 sides, and the left and lower halves were defined as the negative sides.

132

## 133 **Pilot Study**

134           The accuracy of VOG was evaluated in an ideal environment as a pilot study.  
135       Two subjects participated in the pilot study. The target was a rabbit-like character (Fig.  
136       2). The size of the target was 10 × 10 cm, which subtended a visual angle of 5.7° at  
137       1.0 m. The target was displayed on a 24-inch liquid crystal monitor. The center of the  
138       monitor was defined as 0°, the right and upper halves of the monitor were defined as  
139       the positive sides, and the left and lower halves were defined as the negative sides.  
140       The target moved to ±10° with a random velocity of ≤10°/s. The subjects were seated  
141       in a well-lit room (600 lx) wearing their fully corrective spectacles. The subject's head  
142       was fixed with a chin and forehead rest. The subjects were asked to fixate their focus  
143       on the nose of the target whose visual angle was 0.1° at 1.0 m for 60 s.

144

#### 145       **Eye movement test**

146           The main study used the same target with pilot study (Fig 2). The target was  
147       manually moved within ±15° for 60 s by an examiner. All subjects were seated in a  
148       well-lit room (600 lx) wearing their fully corrective spectacles. The subject's head was  
149       fixed with a chin rest and forehead rest. The subjects were asked to fixate on the  
150       nose of the target whose visual angle was 0.1° at 1.0 m during the eye movement  
151       test.

152

### 153 **Training, validation, and test sets**

154           A total of 500 images were extracted from the video recordings of the eye  
155 movement test, which had been recorded by an EMR-9 scene camera for Subject 1.  
156 The training, validation, and test datasets were randomly divided into 300 (60%), 100  
157 (20%), and 100 (20%) images, respectively.

158

### 159 **Object detection algorithm**

160           An SSD was used.<sup>25</sup> The SSD quantifies the output space of the bounding  
161 box into a set of default boxes regardless of the aspect ratio and scale for each  
162 feature map position. The network generates a score for the presence of each object  
163 category in each default box and makes adjustments to the box to better match the  
164 shape of the object in the prediction. The network combines the predictions from  
165 multiple feature maps with different resolutions to handle objects of different sizes.

166           The SSD was divided into the following two parts: shared feedforward  
167 convolutional network and set of subnetworks for classification and regression, which  
168 do not share computations. A pre-trained VGG16 convolutional neural network was  
169 used as a base network.<sup>28, 29</sup> The target images were resized to 300 × 300 pixels in  
170 the preprocessing; then, data augmentation was applied to these target images. The  
171 data that were convolved 10 times were extracted as Source 1 (channel number, 512;

172 feature size,  $38 \times 38$ ) by normalizing the size. The data that were convolved 15 times  
173 were defined as Source 2 (channel number, 512; feature size,  $19 \times 19$ ). The VGG  
174 output of Source 2 was entered into the extra module. The extra module convoluted  
175 Source 2 eight times, and the sources from 3 to 6 (feature sizes,  $10 \times 10$ ,  $5 \times 5$ ,  $3 \times 3$ ,  
176 and  $1 \times 1$ , respectively) were output each time it was convoluted twice. Next, the  
177 sources were entered into the location and confidence module, and these data were  
178 convoluted once. The location module outputted the offset of the default bounding  
179 box. The confidence module outputted the confidence of each class for the default  
180 bounding box.

181 In the training phase, we set the following parameters: 100 epochs, batch  
182 size of 32, and Adam optimizer with the learning rate of 0.001. The validation was  
183 performed every 10 epochs.

184 The SSD model was evaluated by the average precision of the test dataset.  
185 The AP was calculated by the integral of precision and recall. The intersection of  
186 union was calculated by dividing the area of overlap between the predicted bounding  
187 box and the ground-truth bounding box that was included target name and target  
188 location manually defined by an examiner (MH) and the area of union of both  
189 bounding boxes. We defined an IoU  $\geq 75\%$  as "correct." Subsequently, the results of  
190 the test dataset were classified as follows: true positive (TP), predicted bounding box

191 has been covered with ground-truth bounding box ( $IoU \geq 75\%$ ); false positive (FP),  
192 predicted bounding box has been covered with ground-truth bounding box ( $IoU <$   
193  $75\%$  and  $IoU \neq 0$ ); false negative (FN), predicted bounding box has not been  
194 covered with ground-truth bounding box ( $IoU = 0$ ); and true negative, predicted  
195 bounding box and ground-truth bounding box did not exist. Precision was defined as  
196 the percentage at which the IoU can be correctly predicted with an accuracy of  $\geq 75\%$ .  
197 This was calculated by  $TP / (TP + FP)$ . Recall was defined as the percentage at which  
198 the ratio that bounding box at apposition close to the correct answer can be predicted  
199 if the IoU is  $\geq 75\%$  and was calculated by  $TP / (TP + FN)$ . The AP was calculated by  
200 the integral of precision and recall.

201 We used the software Python 3.6.5 on Windows 10 (Microsoft Co., Ltd.,  
202 Redmond, WA, USA) with the following libraries: Matplotlib 3.3.2, Numpy 1.18.5,  
203 OpenCV 3.3.1, Pandas 1.1.3, Pytorch 1.6.0, Scikit-learn 0.23.2, and Seaborn 0.11.0.

204

## 205 **Calculating target location**

206 The SSD drew the location of the object as the bounding box. The bounding  
207 box was computed from two coordinates of  $X_{min}$ ,  $Y_{min}$ , and  $X_{max}$ ,  $Y_{max}$ . The center of  
208 the object coordinates ( $C_x$ ,  $C_y$ ) was determined by  $((X_{max} + X_{min}) / 2, (Y_{max} + Y_{min}) / 2)$ .  
209 Thus, the target location was defined as the center of the bounding box, and a

210 program that outputs the target location in each frame with the recording time  
211 synchronized with the EMR-9 findings to an Excel file (Microsoft Co., Ltd.) in the  
212 inference phase was implemented.

213

## 214 **Data analyses**

215 Data on eye positions and pupil sizes of both eyes were exported to a CSV  
216 file. Data were excluded if the pupil diameter changed by >2 mm/frame due to  
217 blinking.<sup>30</sup> Data were also excluded if the pupil diameter changed by >0.2 mm/frame  
218 over an average of 11 points and median of 5 points due to noise. A linearly  
219 interpolated value replaced the missing values. Horizontal and vertical eye  
220 movements were analyzed, and the SPEM and SEM were identified using a  
221 velocity-threshold identification (I-VT) filter.<sup>31</sup> The I-VT filter was used to classify eye  
222 movements based on the velocity of the directional shifts of the eye. The saccade  
223 was defined as the median velocity of three consecutive windows >100°/s. Then, the  
224 eye position data at 240 Hz were synchronized with the target data at 29.97 Hz.

225 The waveforms of the target locations and eye positions peaked when the  
226 moving direction was reversed (Fig. 3). The latency of SPEM was calculated from the  
227 difference between the horizontal and vertical target location peaks and the dominant  
228 and nondominant eye position peaks. The latency of both eyes was calculated three

229 times or more and averaged for each subject.

230

## 231 **Statistical analyses**

232 The relationships between the target location and both eye positions were  
233 assessed using simple linear regression analysis. The difference between the  
234 latencies of the dominant and nondominant eyes was analyzed by paired *t*-tests. The  
235 relationship between latencies of horizontal and vertical SPEM within both eyes was  
236 assessed using single linear regression analysis.

237 The IBM SPSS Statistics version 26 software (IBM Corp., Armonk, NY, USA)  
238 was used to determine the significance of the differences, and a *P*-value < 0.05 was  
239 considered statistically significant.

240

## 241 **Results**

### 242 **Pilot Study**

#### 243 **Accuracy of VOG in conventional laboratory method**

244 The target location was significantly correlated with both eye positions. The  
245 horizontal and vertical target locations were significantly and positively correlated with  
246 the horizontal dominant (*adjusted R*<sup>2</sup> = 0.989, *P* < 0.001) and nondominant (*adjusted*  
247 *R*<sup>2</sup> = 0.989, *P* < 0.001) eye positions. The vertical dominant (*adjusted R*<sup>2</sup> = 0.987, *P* <

248 0.001) and nondominant ( $adjusted R^2 = 0.987, P < 0.001$ ) eye positions are shown in

249 Figure 4.

250

## 251 **Main study**

### 252 **Demographics**

253           The demographics of the subjects are presented in Table 1. The mean  $\pm$   
254 standard deviation of the refractive errors (spherical equivalents) of the dominant eye  
255 was  $-2.95 \pm 2.46$  diopters (D), and that of the nondominant eye was  $-2.70 \pm 2.60$  D.  
256 The best-corrected visual acuity was 0.0 logMAR units or better in all subjects. The  
257 average heterophoria was  $-1.3 \pm 0.9$  prism diopters (PD) at distance and  
258  $-3.1 \pm 4.4$  PDs at near. All healthy individuals had a stereo acuity of 1.60 log arcsec.

259

### 260 **Integration of deep learning-based object detection and VOG**

261           A representative result for our single shot multibox detector (SSD) is shown in  
262 Movie 1 and Figure 5. The SSD could correctly track a target that an examiner  
263 manually moved. The average precision of the SSD on one class of targets was  
264 97.7% in the test dataset.

265           The horizontal and vertical target locations were significantly and positively  
266 correlated with the horizontal dominant ( $adjusted R^2 = 0.984, P < 0.001$ ) and

267 nondominant (*adjusted R*<sup>2</sup> = 0.983, *P* < 0.001) eye positions. The vertical dominant  
268 (*adjusted R*<sup>2</sup> = 0.955, *P* < 0.001) and nondominant (*adjusted R*<sup>2</sup> = 0.964, *P* < 0.001)  
269 eye positions are shown in Figure 6.

270 The latencies of the horizontal and vertical SPEM were 97.3 ± 26.9 and 115 ±  
271 34.1 ms for the dominant eye (*P* = 0.140), respectively. The latencies of horizontal  
272 and vertical SPEM for the nondominant eye were 108.2 ± 28.6 and 119.1 ± 33.9 ms  
273 (*P* = 0.31), respectively. The latencies of horizontal SPEM were significantly and  
274 positively correlated with the latencies of vertical SPEM in the dominant (*adjusted R*<sup>2</sup>  
275 = 0.824, *P* < 0.001) and nondominant (*adjusted R*<sup>2</sup> = 0.661, *P* = 0.002) eyes.

276

## 277 Discussion

278 Our results show that the combination of VOG and SSD can be used to  
279 evaluate the SPEM of nine directions, and this method can be translated into clinical  
280 settings without changing the testing methods. The SSD recognized the target with  
281 high accuracy, and the target location was significantly and positively correlated with  
282 the positions of both eyes in main study (Figs. 5 and 6). The small variation in values  
283 in the pilot study (Figs. 4 and 6) may be attributable to the fact that the examiner  
284 moved the target by hand, making the SSD's bounding box more susceptible to  
285 deformation because the center coordinates of the bounding box were calculated and

286 converted from pixels to degrees. Nevertheless, the average precision of the SSD on  
287 one class of targets was 97.7% in the test dataset. We considered that SSD has  
288 sufficient accuracy and can be accumulated to evaluate SPEM.

289 Latency or reaction time of SPEM is one of the parameters required to  
290 evaluate strbmismus.<sup>18, 21, 32</sup> In our cohort, the mean latencies of horizontal SPEM  
291 were consistent with those reported by Erkelens and Engel.<sup>33, 34</sup> The latencies of  
292 horizontal and vertical SPEM were not significantly different. Although the testing  
293 methods were different, our findings support those of Rottach et al. who reported that  
294 the latencies of the horizontal and vertical SPEM were not significantly different.<sup>35</sup>  
295 These findings suggest that the current system can accurately determine SPEM in  
296 healthy individuals. Therefore, we will investigate subjects with abnormal SPEM in  
297 further studies.

298 A limitation of our system is the low sampling rate due to the use of a scene  
299 camera. If the target moves at a high speed, the target captured by the scene camera  
300 is blurred, and the accuracy of SSD decreases. Therefore, our system cannot  
301 accurately evaluate saccadic eye movement. In future research, we plan to update  
302 our system to improve the sampling rate of the scene camera to accurately analyze  
303 saccadic eye movement.

304

## 305 **Conclusions**

306           SSD recognized the target that was moved manually with high accuracy, and  
307 the target location was significantly and positively correlated with the positions of both  
308 eyes that were measured by VOG. Our findings suggest that the combination of VOG  
309 and SSD is suitable in evaluating SPEM in the clinic.

310

## 311 **Acknowledgments**

312           All authors thank Mika Suda for illustrating the mascot character. This work  
313 was supported by Grands-in-Aid for Early-Career Scientists, Scientific Research (A)  
314 and (B), Challenging Exploratory Research, Japan Society for the Promotion of  
315 Science, 19K20728 (MH), 18H04116 (MH), 20K04271 (MH), 19K21783 (MH), and  
316 Charitable Trust Fund for Ophthalmic Research in Commemoration of Santen  
317 Pharmaceutical's Founder (MH).

318

## 319 **Author contributions**

320 MH conceived the project and designed the experiments. MH produced the apparatus.  
321 MH performed experiments. MH, TH, , YI, EW, and AM analyzed the data. MH, TH,  
322 and AM wrote the manuscript. All authors reviewed the manuscript.

323

## 324 **Competing interests**

325 All authors declare no competing interests.

326

## 327 **References**

328 1. Rashbass C. The relationship between saccadic and smooth tracking eye  
329 movements. *The Journal of physiology* 1961;159:326-338.

330 2. Robinson DA. The mechanics of human smooth pursuit eye movement. *The*  
331 *Journal of physiology* 1965;180:569-591.

332 3. Westheimer G, McKee SP. Visual acuity in the presence of retinal-image  
333 motion. *J Opt Soc Am* 1975;65:847-850.

334 4. Kowler E, van der Steen J, Tamminga EP, Collewijn H. Voluntary selection of  
335 the target for smooth eye movement in the presence of superimposed, full-field  
336 stationary and moving stimuli. *Vision Res* 1984;24:1789-1798.

337 5. Lisberger SG, Morris EJ, Tychsen L. Visual motion processing and  
338 sensory-motor integration for smooth pursuit eye movements. *Annu Rev Neurosci*  
339 1987;10:97-129.

340 6. Krauzlis RJ. Recasting the smooth pursuit eye movement system. *Journal of*  
341 *neurophysiology* 2004;91:591-603.

342 7. Heinen SJ, Potapchuk E, Watamaniuk SN. A foveal target increases catch-up

343 saccade frequency during smooth pursuit. *Journal of neurophysiology*

344 2016;115:1220-1227.

345 8. Shanidze N, Ghahghaei S, Verghese P. Accuracy of eye position for

346 saccades and smooth pursuit. *Journal of vision* 2016;16:23.

347 9. Helveston EM, Moodley A. How to check eye alignment and movement.

348 *Community Eye Health* 2019;32:55.

349 10. Harley RD. Paralytic strabismus in children. Etiologic incidence and

350 management of the third, fourth, and sixth nerve palsies. *Ophthalmology*

351 1980;87:24-43.

352 11. Helveston EM. Understanding, detecting, and managing strabismus.

353 *Community Eye Health* 2010;23:12-14.

354 12. Vivian AJ, Morris RJ. Diagrammatic Representation of Strabismus. *Eye*

355 1993;7:565-571.

356 13. Pensyl GG, Benjamin WJ. Ocular motility. In: Benjamin WJ (ed), *Borish's*

357 *clinical refraction (2nd ed)*. St Louis: Butterworth Heinemann Elsevier; 2006.

358 14. Levin S, Luebke A, Zee DS, Hain TC, Robinson DA, Holzman PS. Smooth

359 pursuit eye movements in schizophrenics: quantitative measurements with the

360 search-coil technique. *J Psychiatr Res* 1988;22:195-206.

361 15. Imai T, Sekine K, Hattori K, et al. Comparing the accuracy of

362 video-oculography and the scleral search coil system in human eye movement

363 analysis. *Auris Nasus Larynx* 2005;32:3-9.

364 16. Ingster-Moati I, Vaivre-Douret L, Quoc EB, Albuisson E, Dufier JL, Golse B.

365 Vertical and horizontal smooth pursuit eye movements in children: A

366 neuro-developmental study. *Eur J Paediatr Neuro* 2009;13:362-366.

367 17. Bucci MP, Kapoula Z. Binocular coordination of saccades in 7 years old

368 children in single word reading and target fixation. *Vision Res* 2006;46:457-466.

369 18. Lions C, Bui-Quoc E, Wiener-Vacher S, Seassau M, Bucci MP. Smooth

370 pursuit eye movements in children with strabismus and in children with vergence

371 deficits. *PloS one* 2013;8:e83972.

372 19. Lions C, Bui-Quoc E, Seassau M, Bucci MP. Binocular coordination of

373 saccades during reading in strabismic children. *Investigative ophthalmology & visual*

374 *science* 2013;54:620-628.

375 20. Hirota M, Kanda H, Endo T, et al. Relationship between reading performance

376 and saccadic disconjugacy in patients with convergence insufficiency type

377 intermittent exotropia. *Japanese journal of ophthalmology* 2016;60:326-332.

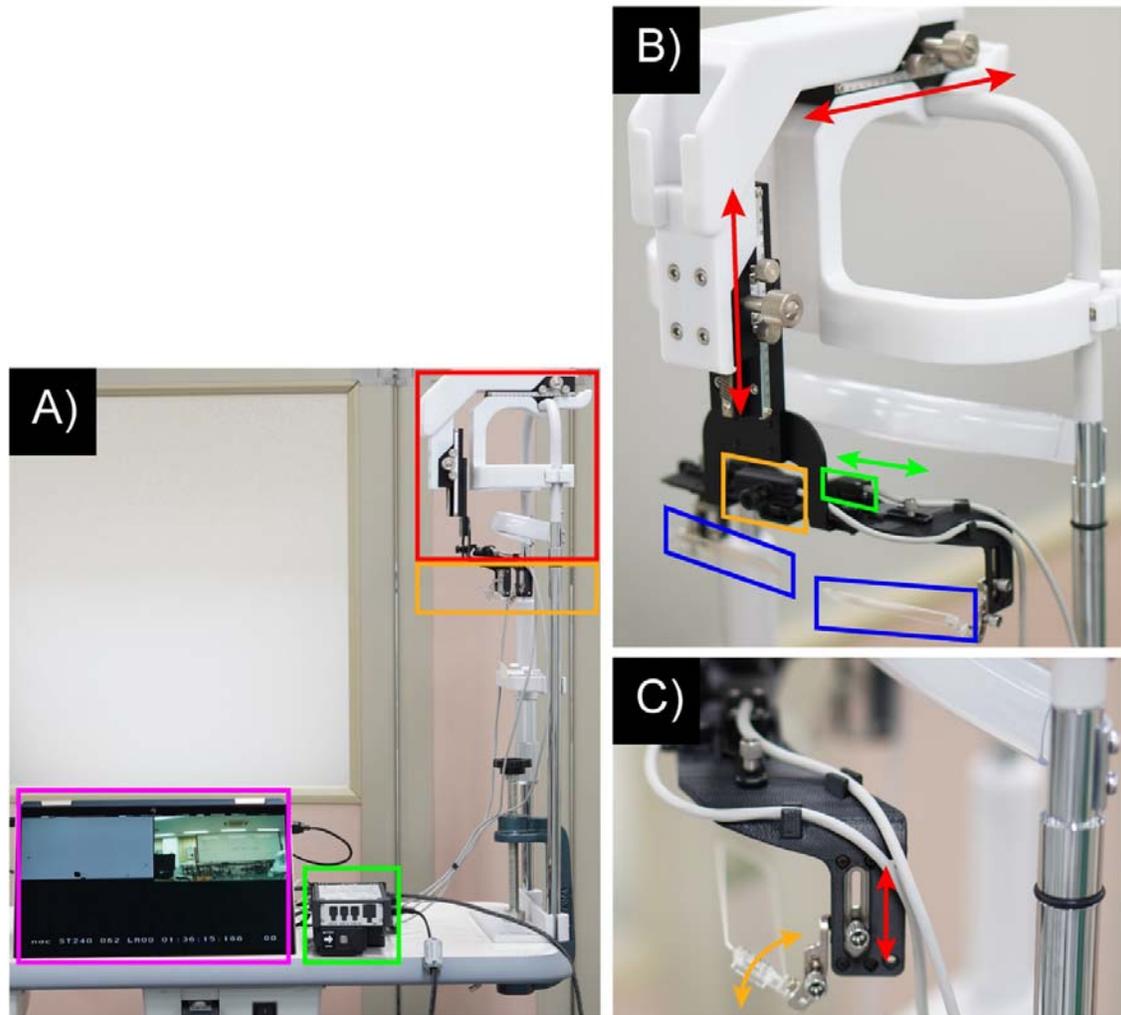
378 21. Fukushima J, Tanaka S, Williams JD, Fukushima K. Voluntary control of

379 saccadic and smooth-pursuit eye movements in children with learning disorders.

380 *Brain Dev-Jpn* 2005;27:579-588.

- 381 22. Metsing I, Ferreira J. The prevalence of poor ocular motilities in a mainstream  
382 school compared to two learning-disabled schools in Johannesburg. *African Vision  
383 and Eye Health* 2015;75:a328.
- 384 23. Raashid RA, Liu IZ, Blakeman A, Goltz HC, Wong AM. The Initiation of  
385 Smooth Pursuit is Delayed in Anisometric Amblyopia. *Investigative ophthalmology  
386 & visual science* 2016;57:1757-1764.
- 387 24. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified,  
388 Real-Time Object Detection. *arXiv e-prints*; 2015:arXiv:1506.02640.
- 389 25. Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector. *arXiv  
390 e-prints*; 2015:arXiv:1512.02325.
- 391 26. Ren SQ, He KM, Girshick R, Sun J. Faster R-CNN: Towards Real-Time  
392 Object Detection with Region Proposal Networks. *Ieee T Pattern Anal*  
393 2017;39:1137-1149.
- 394 27. Cho H, Rybski PE, Bar-Hillel A, Zhang W. Real-time pedestrian detection with  
395 deformable part models. *2012 IEEE Intelligent Vehicles Symposium*  
396 2012;2012:1035-1042.
- 397 28. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition.  
398 *arXiv* 2015;arXiv:1512.03385.
- 399 29. Szegedy C, Vanhoucke V, Loffe S, Shlens J, Wojna Z. Rethinking the

- 400 Inception Architecture for Computer Vision. *arXiv* 2015;arXiv:1512.00567.
- 401 30. Kwon KA, Shipley RJ, Edirisinghe M, et al. High-speed camera  
402 characterization of voluntary eye blinking kinematics. *J R Soc Interface* 2013;10.
- 403 31. Salvucci D, Goldberg J. Identifying fixations and saccades in eye-tracking  
404 protocols. *ETRA '00: Proceedings of the 2000 symposium on Eye tracking research &*  
405 *applications* 2000;71-78.
- 406 32. Kiorpes L, Walton PJ, O'Keefe LP, Movshon JA, Lisberger SG. Effects of  
407 early-onset artificial strabismus on pursuit eye movements and on neuronal  
408 responses in area MT of macaque monkeys. *The Journal of neuroscience : the official*  
409 *journal of the Society for Neuroscience* 1996;16:6537-6553.
- 410 33. Engel KC, Anderson JH, Soechting JF. Oculomotor tracking in two  
411 dimensions. *Journal of neurophysiology* 1999;81:1597-1602.
- 412 34. Erkelens CJ. Coordination of smooth pursuit and saccades. *Vision Research*  
413 2006;46:163-170.
- 414 35. Rottach KG, Zivotofsky AZ, Das VE, et al. Comparison of horizontal, vertical  
415 and diagonal smooth pursuit eye movements in normal human subjects. *Vision Res*  
416 1996;36:2189-2195.
- 417



418

419 **Figure 1. Custom-made eye tracker based on EMR-9**

420 The exterior of the custom-made eye tracker was based on EMR-9 (A). The total area

421 of the red and orange squares indicates EMR-9, and the red and orange squares

422 indicate (B) and (C). The green square indicates the controller of the EMR-9. The

423 purple square indicates the output image by EMR-9.

424 In (B), the orange, green, and blue squares indicate the scene camera, eye camera

425 (there is also one for the right eye on the other side), and half-mirrors, respectively.

426 The scene camera is able to rotate 60° in pitch. Both eye cameras can horizontally

427 shift to 1.3 cm (total, 2.6 cm) to adjust the pupillary distance (green two-direction  
428 arrow line). The base position of the EMR-9 can adjust 8.0 cm horizontally and  
429 vertically (red two-direction arrow lines).

430 In (C), the half-mirror can shift to 2.5 cm vertically (red two-direction arrow line) and  
431 rotate 30° in pitch (orange two-direction arrow line).

432

433

434

435

436

437

438

439

440



441

442 **Figure 2. Examination target**

443 The rabbit-like character is a mascot of the Department of Orthoptics, Teikyo

444 University (created by Mika Suda). The subjects were asked to fixate on the nose of

445 the target whose visual angle was  $0.1^\circ$  at 1.0 m during the eye movement tests.

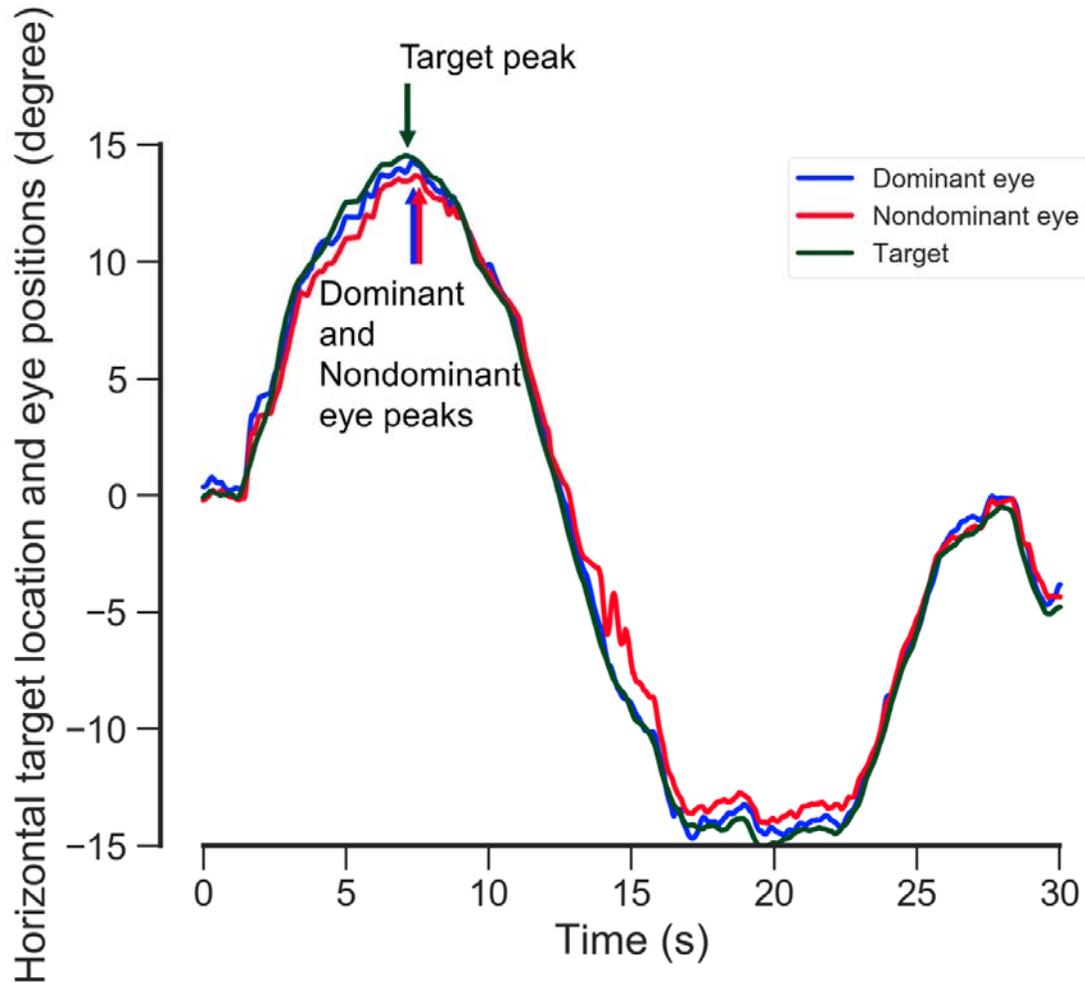
446

447

448

449

450



451

452 **Figure 3. Graph showing the latencies of SPEM for the dominant and**

453 **nondominant eyes**

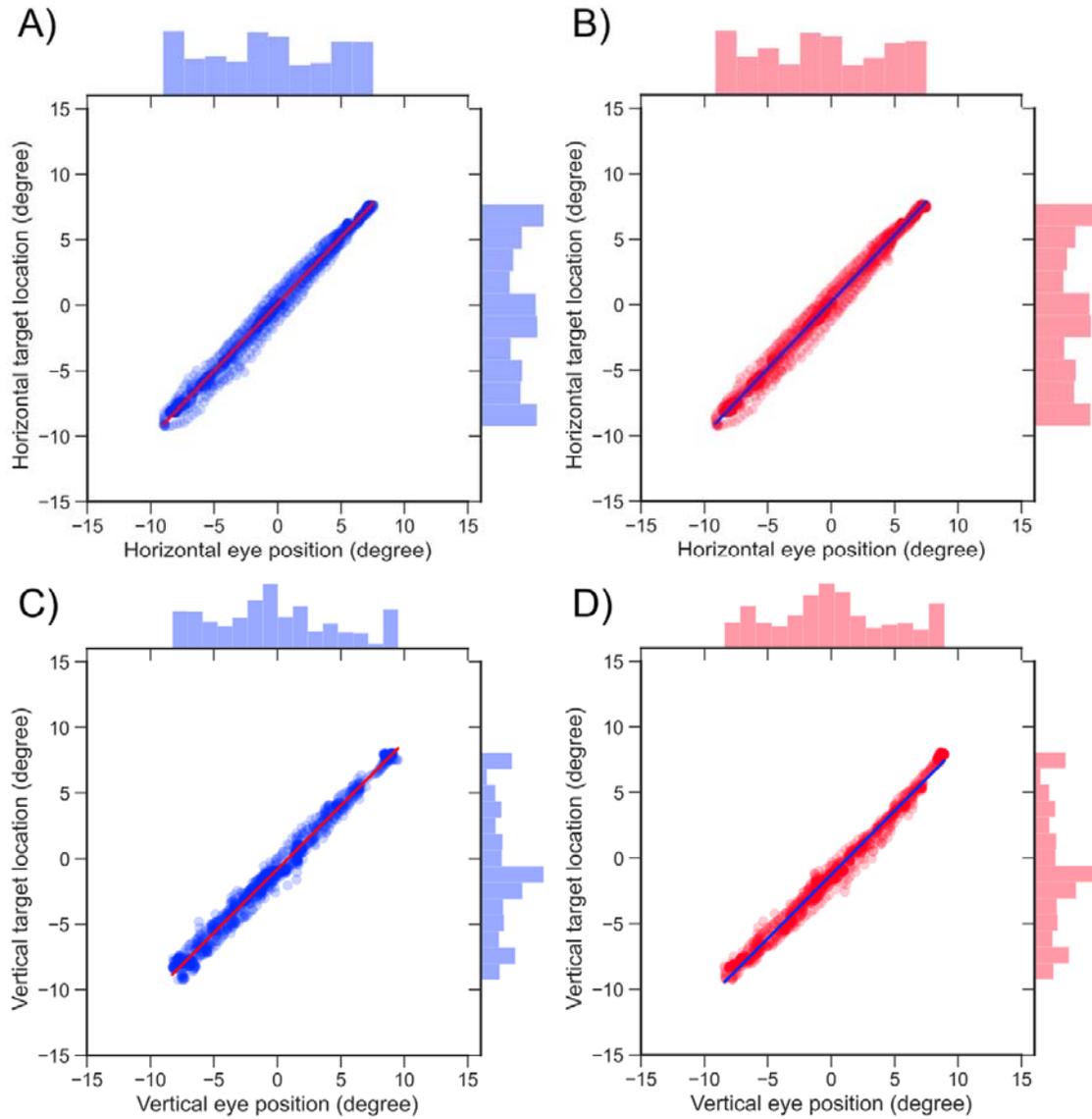
454 The blue, red, and green lines indicate the horizontal dominant and nondominant eye

455 positions and target location during the eye movement test. The latencies of SPEM

456 were calculated from the difference between the target location peak and the

457 dominant and nondominant eye position peaks.

458



459

460 **Figure 4. Correlations between horizontal (A, B) and vertical (C, D) target**

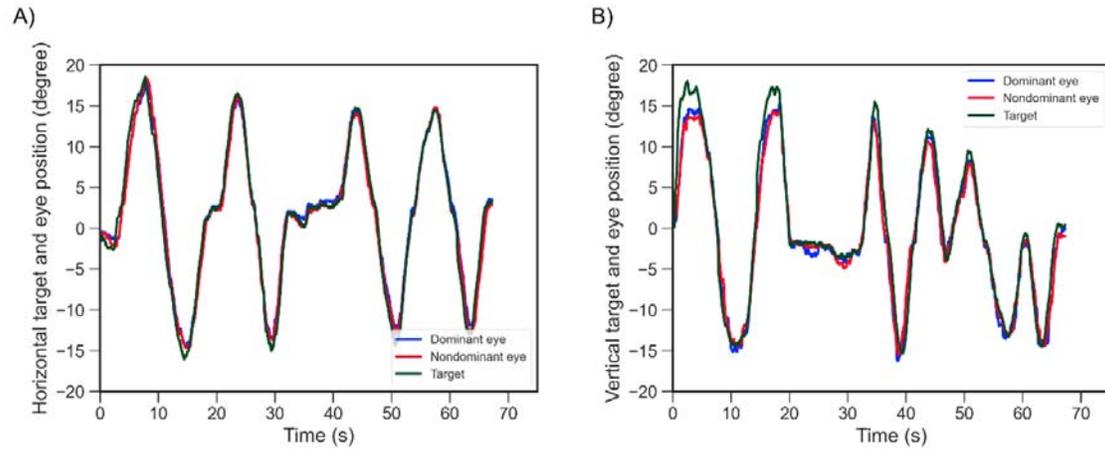
461 **locations and eye positions in a pilot study (n = 2)**

462 The blue and red dots indicate the relationships between the target and dominant or

463 nondominant eye velocity. The red and blue lines indicate the regression lines. The

464 histograms at the upper and right sides indicate the distribution between the target

465 location and the dominant or nondominant eye position.



466

467 **Figure 5. Horizontal (A) and vertical (B) target locations and eye positions when**

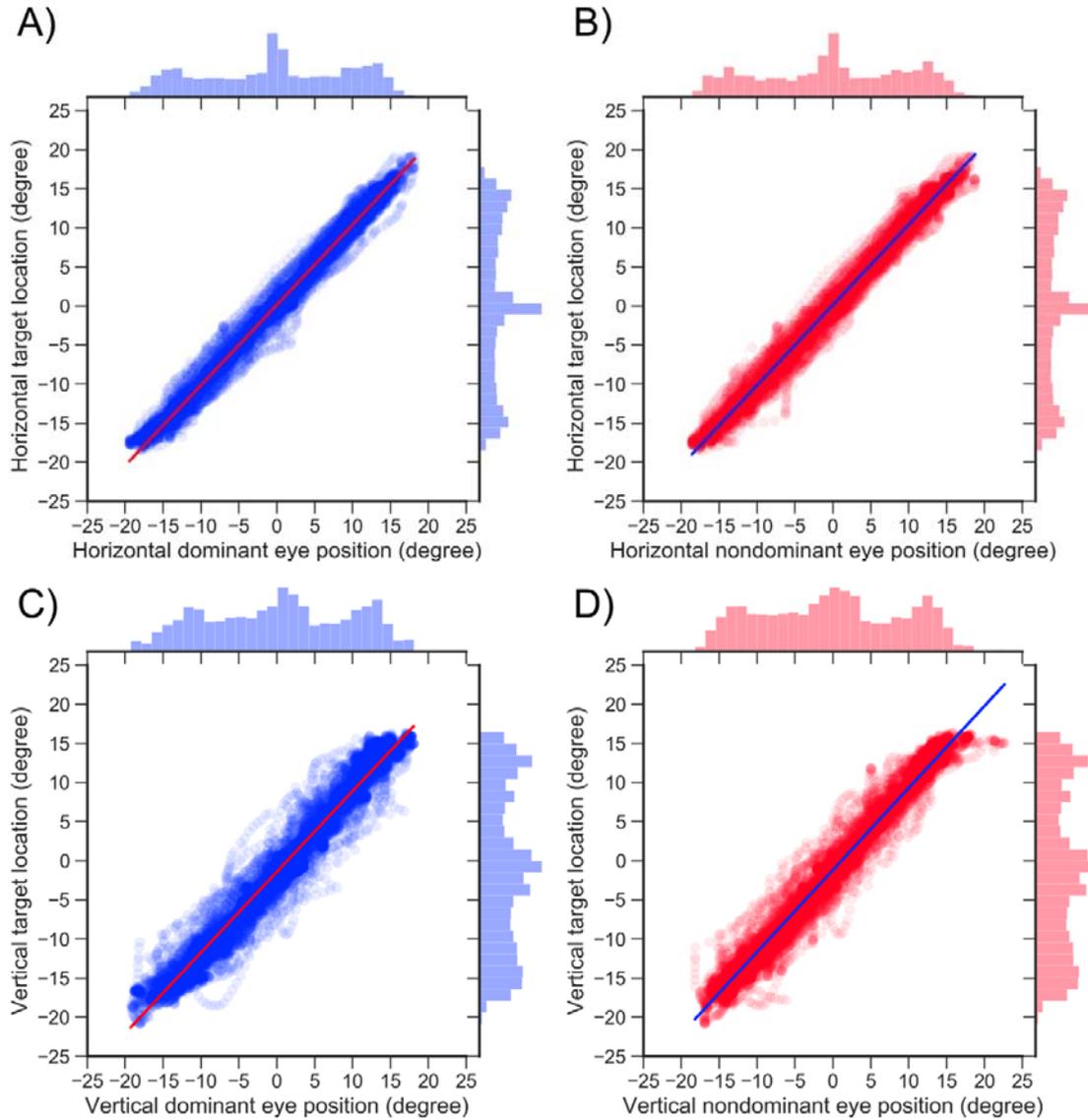
468 **the examiner moved the target by hand.**

469 The blue, red, and green lines indicate the horizontal dominant and nondominant eye

470 positions and target location during the eye movement test.

471

472



473

474 **Figure 6. Correlations between horizontal (A, B) and vertical (C, D) target**

475 **locations and eye positions**

476 The blue and red dots indicate the relationships between the target and the dominant

477 or nondominant eye velocity. The red and blue lines indicate the regression lines. The

478 histograms at the upper and right sides indicate the distribution between the target

479 location and the dominant or nondominant eye position.