

1 **TITLE:** Automatic Processing of Gaze Movements to Quantify Gaze Scanning Behaviors in a
2 Driving Simulator.

3 **AUTHORS:** Garrett Swan^a, Robert B. Goldstein^a, Steven W Savage^a, Lily Zhang^a, Aliakbar
4 Ahmadi^{a,b}, Alex R Bowers^a

5

6 **AUTHORS' INSTITUTION:**

7 ^aSchepens Eye Research Institute of Massachusetts Eye and Ear, Department of Ophthalmology,
8 Harvard Medical School

9 ^bTechnical University of Munich, Department of Mechanical Engineering,

10

11 **GRANT:** NIH grants R01-EY025677, P30-EY003790, and S10-RR028122 and with support from
12 the Hand-Böckerler-Foundation

13

14 **Corresponding author:**

15 Garrett Swan

16 20 Staniford St

17 Boston, MA, 02114

18 919-221-5345

19 gsp.swan@gmail.com

20

21

22 **ABSTRACT**

23 Eye and head movements are used to scan the environment when driving. In particular,
24 when approaching an intersection, large gaze scans to the left and right, comprising of head
25 and multiple eye movements, are made. We detail an algorithm called the gaze scan algorithm
26 that automatically quantifies the magnitude, duration, and composition of such large lateral
27 gaze scans. The algorithm works by first detecting lateral saccades, then merging these lateral
28 saccades into gaze scans, with the start and end point of each gaze scan marked in time and
29 eccentricity. We evaluated the algorithm by comparing gaze scans generated by the algorithm
30 to manually-marked ‘consensus ground truth’ gaze scans taken from gaze data collected in a
31 high-fidelity driving simulator. We found that the gaze scan algorithm successfully marked 96%
32 of gaze scans, produced magnitudes and durations close to ground truth, and the differences
33 between the algorithm and ground truth were similar to the differences found between expert
34 coders. Therefore, the algorithm may be used in lieu of manual marking of gaze data,
35 significantly accelerating the time consuming marking of gaze movement data in driving
36 simulator studies. The algorithm also complements existing eye tracking and mobility research
37 by quantifying the number, direction, magnitude and timing of gaze scans and can be used to
38 better understand how individuals scan their environment.

39

40 **Keywords:** gaze tracking, eye and head scanning, eye movement event detection, driving
41 simulation

42

43 **1. Introduction**

44 When driving we use head and eye movements to scan the environment to search for
45 potential hazards and to navigate. Scanning is especially important when approaching
46 intersections, where a large field of view (e.g., 180° at a T-intersection) needs to be checked for
47 vehicles, pedestrians, and other road users. Typically, drivers make left and right scans that
48 start near and return to the straight ahead position. The scans become increasingly larger in
49 magnitude as the driver approaches an intersection with larger scans requiring different
50 numbers and sizes of eye and head movements (Figure 1). Insufficient scanning has been
51 suggested as one mechanism for increased crash risk at intersections (Hakamies-Blomqvist,
52 1993). Previous studies have reported that older adults scan insufficiently at intersections
53 compared to younger adults in on-road driving (Bao & Boyle, 2009a) and in a driving simulator
54 (Romoser & Fisher, 2009; Romoser, Pollatsek, Fisher, & Williams, 2013; Savage et al., 2017;
55 Bowers et al., 2019; Savage et al., *Revise and Resubmit*). Individuals with vision loss have also
56 been found to demonstrate scanning deficits at intersections in a driving simulator (Bowers,
57 Ananyev, Mandel, Goldstein, & Peli, 2014). These studies and analyses of police crash reports
58 (McKnight & McKnight, 2003; Braitman, Kirley, McCartt, & Chaundhry, 2008) suggest that
59 scanning plays an important role in driving and that quantifying scanning may provide insights
60 into why some individuals fail to detect hazards at intersections. Here, we are interested in
61 quantifying visual scanning as lateral gaze scans, which encompass all of the gaze movements
62 (the combination of eye and head movements) that extend horizontally from the starting point
63 near the straight ahead positon to the maximally eccentric gaze position. This research extends
64 our previous quantification of head scans (Bowers et al., 2014) by taking account of eye
65 position as well as head position to characterize gaze scanning while driving.

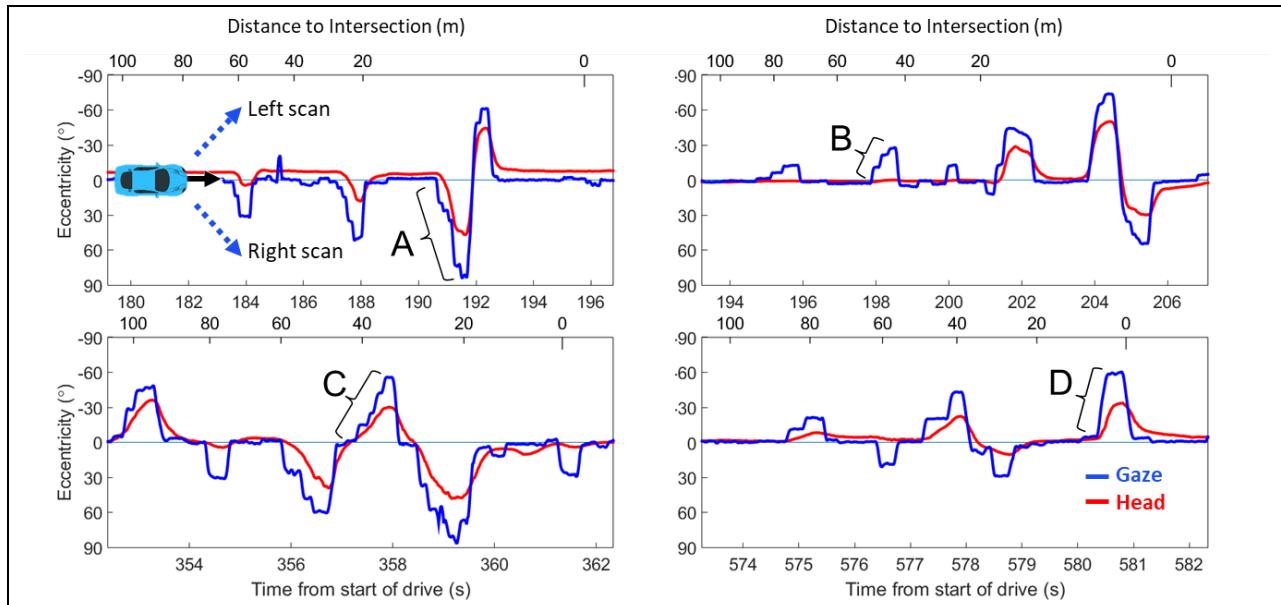


Figure 1. Examples of the diversity of individuals' scanning patterns on approach to an intersection (gaze = blue, head = red). Sections of these plots will be used in subsequent figures to illustrate different aspects of the gaze scan algorithm. Each plot shows data from 100 to 0 m before the intersection. The black arrow in front of the car in the top left plot indicates the travel direction (i.e. left to right means forward in time). Participants decelerated at different rates, hence the different spacings between tick marks on the top (distance-to-intersection) axis. The dotted blue arrows in the top left plot indicate the direction of the gaze and head scans. Any scan below 0° eccentricity is a scan to the left and any scan above 0° eccentricity is a scan to the right. Some gaze scans were made with large head movements (e.g. A), while others were made without any head movements (e.g. B). Some large (60°) scans were slow and comprised of multiple saccades (e.g., C) while others were quick and comprised of only one saccade (e.g., D).

67

68 Studies have used different techniques to combine eye and head tracking when driving
 69 to better understand how drivers scan while approaching an intersection. One approach is to
 70 quantify the standard deviation of the horizontal displacements in gaze to capture effects such
 71 as visual tunneling, or the lack of looking into the periphery (Sodhi, Reimer, & Llamazares, 2002;
 72 Reimer 2009). One limitation of this approach is that it does not quantify how many times

73 someone scanned to the left or right nor does it provide information about the gaze
74 movements that compose the scan. Some studies have quantified scanning by manually
75 counting discrete head turns while participants were driving (Keskinen, Ota, & Katila, 1998;
76 Romoser & Fisher, 2009; Bao & Boyle, 2009b; Romoser, Fisher, Mourant, Wachtel, & Sizov,
77 2005). However, categorizing scans as only ‘left’ or ‘right’ fails to capture the magnitude of
78 those scans and how those scans were made (i.e. the composition of head and eye
79 movements). Other studies have quantified scanning by overlaying eye position onto video of
80 the driving scene to manually determine the location of lateral gaze movements (Romoser,
81 Pollatsek, Fisher, & Williams, 2013) or by manually marking the start and end of lateral gaze
82 movements (Alberti, Goldstein, Peli, & Bowers, 2017). While manual marking of gaze
83 movements is common in the literature, it is extremely time consuming, especially when the
84 individual doing the marking must look through video frame by frame, and could be prone to
85 potential inconsistencies when there are multiple individuals marking gaze movements. An
86 alternative to manual marking is automatic detection of gaze movements using an algorithm,
87 which could mark eye and head movements in lieu of manual marking altogether. The
88 algorithm could also be used to parse data into simpler chunks for expert coders (Munn,
89 Stefano, & Pelz, 2008).

90 Bowers and colleagues (2014) created an algorithm that automatically quantified the
91 magnitude, direction, and numbers of lateral *head scans* on approach to intersections. That
92 algorithm detected large discrete rotations that took *head eccentricity* at least 4° away from
93 the straight ahead position for at least 0.2 s. While that algorithm successfully marked large
94 lateral head movements, it did not account for eye position. To fully understand scanning
95 behaviors when driving, we need to be able to quantify gaze movements, which are the
96 combination of head-in-world and eye-in-head movements. Gaze movements differ from head
97 movements in driving: they tend to have faster velocities, extend further laterally, and are often
98 composed of multiple discrete saccades and fixations that resemble staircases (e.g., scan C in
99 Figure 1). Given the differences between gaze and head movements when scanning, the head
100 movement detection algorithm (Bowers et al., 2014) is not suitable for marking lateral gaze
101 movements.

102 Alternatively, one could utilize eye tracking event detection algorithms (e.g. Salvucci &
103 Goldberg, 2000; Nyström & Holmqvist, 2010) that detect fixations and saccadic eye
104 movements. However, these algorithms are not appropriate by themselves for detecting gaze
105 events for two reasons. Firstly, gaze movements that exceed the typical oculomotor range
106 ($\pm 50^\circ$) are slower than smaller gaze movements given that at least part of the gaze movement
107 must be composed of head rotation (Barnes, 1979; Guitton & Volle, 1987). Therefore, the
108 parameters for detecting saccades from gaze will likely differ from the parameters typically
109 used for detecting eye-only saccades. Secondly, event detection may capture the eye
110 movements that compose a gaze scan, but additional steps would be required for these
111 markings to be interpretable for large gaze scans. For example, to know how far an individual
112 looked, which may be a gaze scan composed of multiple saccades, one would need to
113 determine from the series of saccades which was the most eccentric, requiring additional
114 computation beyond simply marking each saccade. Therefore, we define and measure gaze
115 scans as the entire horizontal movement of the eyes plus head that can be composed of one or
116 more saccades. Here we present an algorithm called the gaze scan algorithm that automatically
117 marks gaze scans by merging neighboring saccades into a single gaze scan that ends at the most
118 eccentric gaze location.

119 The goal of this algorithm is to mark the start and end of each gaze scan in time and
120 eccentricity in order to quantify the direction, timing, magnitude, and composition of the gaze
121 scan. Our approach to marking gaze scans is reductionist: first, we take a subset (bracketing a
122 known event or section of road) of gaze data, isolate saccades, and then merge those saccades
123 into gaze scans. This approach has several advantages: 1) it is based on gaze movements and
124 not head movements which is important because not all gaze movements have a head
125 component (see Figure 1; Savage et al., *Revise and Resubmit*), 2) the merging of saccades is
126 independent of sampling rate and can be paired with any event-detection algorithm, 3)
127 provides information about the saccades that compose the gaze scans, and 4) can be used to
128 quantify the number of gaze scans, regardless of the magnitude or duration of the gaze scan. In
129 order to develop and evaluate this algorithm, the algorithm's marking of gaze scans was
130 compared to manually marked gaze scans from data collected while participants drove in a

131 high-fidelity driving simulator. A successful outcome would enable much more efficient
132 processing of gaze data in future driving simulator studies.

133 **2. Materials and Methods**

134 **2.1 Participants**

135 The gaze scan algorithm was evaluated using data from a previous study (Savage et al.,
136 *Revise and Resubmit*), approved by the institutional review board at the Schepens Eye Research
137 Institute. Given the large number of scans in the original data set and the time consuming
138 nature of manual marking, a subset of the data were used in the evaluation of the algorithm.
139 Data were pseudo-randomly selected from the original dataset to ensure a mix of gender and
140 age in the sample. In total 19 drives from 13 unique participants out of the original 29
141 participants were selected. These 13 participants had been recruited from local advertisements
142 (IRB-approved) and from a database of participants who had participated in previous studies or
143 were interested in participation. They were current drivers with at least two years of driving
144 experience, average binocular visual acuity of 20/20, and no self-reported adverse ocular
145 history. Six of the 19 drives were from female drivers and six of the drives were from older
146 drivers (+65 years old) compared to those from younger drivers (20-40 years old), which are
147 similar to the proportion of demographics in Savage et al. The data from these 19 drives were
148 split into data sets that are described further in section 3.2.

149 **2.2 Apparatus**

150 The driving simulator (LE-1500, FAAC Corp, Ann Arbor, MI) presented a virtual world at
151 30 Hz onto five, 42-inch liquid-crystal display (LCD) monitors (LG M4212C-BA, native resolution
152 of 1366 x 768 pixels per monitor; LG Electronics, Seoul, South Korea) that offered
153 approximately 225° horizontal field of view of the virtual world (Figure 2). The simulator was
154 fully controlled by the participant in a cab, which included a steering wheel, gear shifter, air
155 conditioning, turn signal, rear and side mirrors (inset on the monitors), speedometer (inset on
156 the central monitor), and a motion seat. The virtual environment was created with Scenario
157 Toolbox software (version 3.9.4. 25873, FAAC Incorporated) and was set in a light industrial

158 virtual world consisting of an urban environment with roads set out on a grid system with many
159 four-way (+) and three-way (T) intersections. The world contained a variety of buildings, other
160 traffic on the road, and signage (e.g. stop signs, traffic lights). All participants drove the same
161 route through 42 intersections and approximately half of these intersections included cross-
162 traffic that appeared on the left, right, or straight ahead (see Savage et al. for details).



Figure 2. Image of the driving simulator equipped with 6 cameras (red circles) located around the driver's seat (two on the left, two on the right, and two in the center), which enabled recording of lateral eye and head position up to 90° to the left and right of the driver.

163

164 While driving in the virtual world, head and eye movements were tracked across 180°
165 (90° to the left and right of the straight ahead position), which is sufficient for capturing large
166 lateral eye and head scans on approach to intersections. Eye and head positions were recorded
167 at 60 Hz with a remote, digital 6-camera tracking system (Smart Eye Pro Version 6.1, Goteborg,
168 Sweden, 2015) located around the participant (see Figure 2, red circles). Gaze tracking was
169 achieved using the pupil corneal reflection and estimating the combined position and direction
170 of a 3D profile of both eyes. Head tracking was achieved automatically by creating a 3D profile
171 of the participant's face using salient features (e.g. eye corners, nostrils, mouth corners, and
172 ears) to capture the position and direction of the head. Following data collection, the eye and
173 heading tracking data and the driving simulator data were synchronized via time stamps.

174 2.3 Procedure

175 Participants drove through an acclimatization drive and practice drive (approximately 8
176 to 10 minutes each) to become familiar with driving the simulator. Participants were instructed
177 to drive (speed capped at 35 mph) as they would in the real world, obey traffic rules, and press
178 the horn whenever they saw a motorcycle (included motorcycle hazards approaching from a
179 cross road at 16 intersections). Participants were not given any instructions regarding how or
180 when to scan. Prior to the experimental drives, each camera's position was adjusted
181 sequentially to capture as much of the face as possible in the camera's field of view, followed
182 by any necessary adjustments to the aperture and focus. The cameras were calibrated with a
183 checkerboard pattern that was presented to each camera from the location of the driver's
184 head. The head position was tracked automatically after camera calibration by detecting
185 features of the participant's face. The eyes were calibrated with 5 points on the center screen
186 in the driving simulator. Verification of the calibrations resulted in a median accuracy of 2.6°
187 and precision of 1.6° for the 5 calibration points. In each of the two experimental drives,
188 participants drove through 42 pre-determined intersections in the same virtual city. For the
189 purposes of this paper, we only considered data that corresponded to 100 m before and up to
190 the white line at T and + intersections (total of 32 intersections per drive, half of which
191 contained hazards). For a full description of the procedure, see Savage et al. (*Revise and*
192 *Resubmit*).

193 2.4 Post processing

194 Following data collection, data were processed in MATLAB (Mathworks, R2015a). Eye
195 movement data are typically contaminated with data loss (i.e., loss of tracking, or sections
196 where the eyes could not be tracked) and noise. To remove these irregularities, we
197 implemented an aggressive outlier removal process using two sequential all-zero (finite impulse
198 response: FIR) filters (we used the Matlab function *filtfilt.m* with window sizes of 33 and 66 ms
199 respectively). Median filtering was chosen because it does not alter any data and preserves high
200 frequency events. We first removed large outliers and then smaller ones by removing data
201 points that differed by 16° between raw and filtered. Sometimes neighboring points were
202 influenced by large outliers, so we repeated this step using a threshold of 8°. We then removed

any remaining data points with velocities that exceeded the physical limits of eye movements. Unphysical velocities were defined as velocities that exceeded thresholds from the main sequence as described in Bahill et al. (1975), given an assumed fixed relationship between saccade magnitude and peak velocity. These processing steps were applied to all data and data points that were missing due to loss of tracking or removed because of noise and were replaced using a linear interpolation. The 60 Hz data were then smoothed with a Savitzky-Golay filter (*sgolayfilt.m*, with filter order = 3, filter length = 0.117 ms [7 samples]) to preserve high-frequency peaks (Savitzky & Golay, 1964; Nyström & Holmqvist, 2010). Post-processed data were used during manual marking and for processing gaze scans using the gaze scan algorithm.

3. Gaze scan algorithm

3.1. Defining a gaze scan

When approaching an intersection, gaze movements typically start from and return to close to the straight ahead position (Figure 1). We therefore define a gaze scan as any lateral gaze movement that takes the eyes away from the straight ahead position (i.e. 0°) into the periphery. Gaze scans could be composed of a single or multiple saccadic gaze movements (e.g. see Figure 1) and were always defined as the whole movement from the starting point near straight ahead to the maximum eccentricity towards the left (defined as gaze scans between 0° and -90° eccentricity) or right (gaze scans between 0° and 90° eccentricity). Gaze movements that returned to 0°, which we define as return gaze scans were not analyzed here because it is only the scans headed away from the straight ahead position that capture the extent of lateral scans. In some instances, the return gaze scan did not stop at the straight ahead position, but continued to the opposite side. Any such gaze scans that crossed the straight ahead position (0°) were split into one return and one away gaze scan (see section 3.3.2). Thus, gaze scans contain side (i.e. on the left or right side of 0°) and direction (i.e. towards the left or right side of 0°) information. Each gaze scan has a start and end time and eccentricity. The duration of a gaze scan was calculated as the difference in time between the start and end of the gaze scan. The magnitude of a gaze scan was calculated as the difference in eccentricity between the start and end of the gaze scan. Given this information, other variables could be defined with respect

231 to the timing of a gaze scan, such as the size of the head movement component of the gaze
 232 scan, or the speed and distance of the car to the intersection at the time of the start of the gaze
 233 scan.

234 3.2 Manually marked gaze scans

235 Three authors (G.S., S.W S., and L.Z.) manually marked gaze data from the 19 selected
 236 drives which were randomly split between two sets of data (see Table 1). The first set (ground-
 237 truth data set) was used to optimize and evaluate the gaze scan algorithm and contained
 238 manually marked gaze scans that the three expert coders agreed upon (i.e., consensus marking
 239 with all three coders in the same room viewing the same monitor). This set was further split
 240 pseudo-randomly by drive (i.e., total driving route) into a training set for the optimization of the
 241 gaze scan algorithm, and a testing set for the evaluation of the gaze scan algorithm. The second
 242 set (coders' data set) was used to quantify the variance in marking between the three expert
 243 manual coders and contained manually marked gaze scans that the three expert coders marked
 244 individually. A total of 4246 gaze scans were marked, which corresponded to 6873 seconds of
 245 driving data (see Table 1 for details).

	Type of manual marking	Number of drives	Number of gaze scans	Duration of driving data (seconds)	Purpose
Ground Truth data Training set	Consensus	8	2322	3461	Optimize gaze scan algorithm merging parameter
Ground Truth data Testing set	Consensus	4	1094	1861	Evaluation of gaze scan algorithm performance
Coders' data	Individual	7	830	1551	Estimate expected variance between coders when manual marking
Table 1. Details of Ground-truth and Coders' data sets					

247 *Methods for manual marking:* Using the post-processed data, the three expert coders
248 manually marked gaze scans headed away from the straight ahead position using a custom
249 MATLAB GUI that presented lateral gaze and head eccentricity and the time the driver entered
250 an intersection. Manual coders marked gaze scans from subsets of the data that corresponded
251 to when the driver was approximately 100 m before and up to the time the driver entered the
252 intersection (crossed the white line of each intersection), which resulted in approximately 13.5
253 seconds (st.d. = 3 seconds) of data being presented at a time on the x-axis. The y-axis range was
254 the same on all plots, set from -90° to 90° to capture all possible horizontal gaze movements.
255 This format was exactly the same as the presentations in Figure 1. The three expert coders
256 marked gaze scans sequentially by selecting the eccentricity and time a gaze scan started and
257 then ended according to our definition of a gaze scan (section 3.1). This was achieved by
258 clicking on the graph twice (first for the start and second for the end of a gaze scan), and then
259 clicking a third button that connected the two points to create a gaze scan. Only gaze scans

260 *Gaze scan matching:* We developed a procedure to match gaze scans. This procedure
261 was used to match the algorithm to the ground truth and to match gaze scans between two
262 different coders. The below description thus matches gaze scans from set B (e.g., algorithm) to
263 set A (e.g., ground truth). Matching was done based on the scan start time, end time, and the
264 midpoint between the start and end times.

265 For a given scan in set A, we searched all of set B's scans for those with a midpoint
266 between the start and end time of the given scan in set A. We also searched all of set B's scans
267 for those with a start and end time that contained the midpoint for the given scan in set A.

268 If the initial searches returned a single scan from set B, we next checked if the start and
269 end time of that scan in set B contained the midpoint of multiple scans from set A. If so, then
270 those scans were paired with the single scan in a many-to-one match (section 3.4). Otherwise it
271 was designated as a one-to-one match. If the initial searches returned multiple scans from set
272 B, then those scans were paired with the given scan in set A as a one-to-many match (section
273 3.4).

274 When matching scans from set B to set A, the procedure only included those scans from
275 set A that had no prior matching scans to set B. That is, for a scan in set A already paired in a
276 many-to-one match, that scan did not go the matching procedure again. The matching
277 procedure may return some scans in set A and set B with no matches.

278 *Ground-truth data set:* For each manually-marked gaze scan, consensus between the
279 three expert coders was required before accepting the gaze scan to be part of the ground-truth
280 data set. Consensus was achieved by having all three expert coders view the same image
281 simultaneously and having at least two out of three coders agree on the start and end of each
282 gaze scan. The scans from the algorithm that could not be matched with any ground truth scan,
283 and vice versa, were omitted from analyses. Only a small percentage of the ground truth scans
284 were omitted from analyses (testing set = 2.0%) with the majority of these being cases where
285 the algorithm did not mark the gaze data as being a saccade (testing set = 75.9%) or cases
286 where the algorithm and manual marking were offset in time and thereby not properly paired
287 (testing set = 24.1%).

288 *Coders' data set:* These gaze data were independently marked by the three expert
289 coders and then used to quantify the level of agreement amongst them in their manual
290 markings. This provided a comparison for the level of agreement between the gaze scan
291 algorithm and the ground-truth testing set. In the coders' data set, approximately 16% of the
292 gaze scans were omitted from our analyses because there was no matching gaze scan from
293 either of the other manual coders.

294 3.3 Gaze scan algorithm implementation

295 The gaze scan algorithm was implemented in MATLAB (Mathworks, R2015a). The gaze
296 scan algorithm automatically marked gaze scans in two stages. First, gaze data were reduced to
297 saccades (defined in next section, 3.3.1). The second stage of the gaze scan algorithm was to
298 merge the sequences of saccades into gaze scans based on a set of rules. A detailed diagram for
299 how the algorithm processes data is provided in the Appendix (section A.1). Furthermore, code
300 for the gaze scan algorithm and manual marking can be downloaded from <https://osf.io/p6jqn/>.

301 3.3.1 First stage of Gaze scan algorithm – Saccade detection

302 Saccades (Figure A.3 in Appendix A.3) were found by calculating the velocity between
303 each gaze sample using the smoothed eccentricity and time. If two points had a velocity greater
304 than 30 °/s, then both samples were marked as belonging to a saccade. To capture onset and
305 offset velocities of saccades, we opted for a velocity threshold below what is typically used for
306 detecting eye saccades (e.g. 75 °/s; Smeets & Hooge, 2003), given that large saccades that have
307 a head movement component may have slower velocities than eye saccades without any head
308 movement component (Barnes, 1979; Guitton & Volle, 1987). A similar 30 °/s velocity threshold
309 for detection of saccades has also been used in other studies involving driving simulation and
310 gaze tracking (e.g. Hamel et al., 2013; Bahnemann, et al., 2015). Only neighboring data points
311 that exceeded the velocity threshold and were headed in the same direction were combined to
312 form a saccade. The onset and offset of a saccade was defined by the first and last data point.
313 Saccades that had a lateral magnitude smaller than 1° or were shorter than 2 samples (0.033 s)
314 were removed in order to minimize the likelihood of marking noise as a saccade (see Beintema,
315 Van Loon, & Van Den Berg, 2005 for a similar approach).

316

317 3.3.2 Crossing zero line

318 While the majority of the gaze scans start and end near the straight ahead position (0°),
319 some saccades from gaze scans cross 0°. Saccades that crossed the straight ahead position were
320 split into two saccades (Figure 3). By splitting saccades with respect to the straight ahead
321 position, we can directly compare left and right gaze behavior with objects that appear on the
322 left and right in the environment. Furthermore, in a post-hoc analysis, over 70% of the gaze
323 scans started within 7° of the straight ahead position. When splitting saccades that cross 0°, the
324 new first saccade now contained a linearly interpolated gaze and time value immediately
325 before the cross over, while the new second saccade now contained the value immediately
326 after the cross over. Because the saccade was split into two new saccades, it necessitated that
327 the two new saccades still satisfied the thresholds for saccade detection (section 3.3.1). Any

328 new saccade created after splitting two saccades that no longer satisfied the rules was no
 329 longer categorized as a saccade.

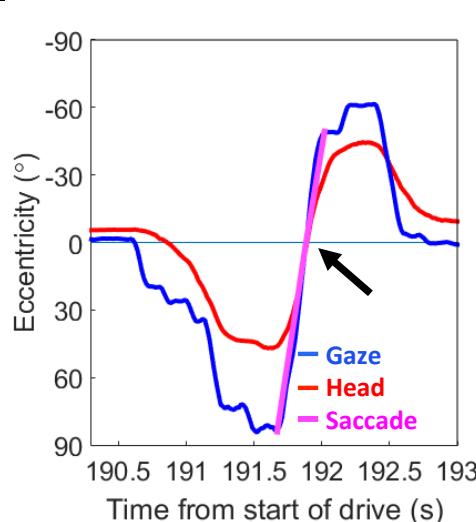


Figure 3. Zoomed in data from Figure 1 (upper left plot), illustrating where a saccade is split when crossing 0°.

330

331 3.3.3 Second stage of gaze scan algorithm – merging saccades into scans

332 The sequence of saccades was next merged into gaze scans (Figure 4). Any two saccades
 333 could be merged to form a gaze scan headed away from 0°. Merging occurred by comparing
 334 two saccades and merging those saccades if they satisfied the following rules:

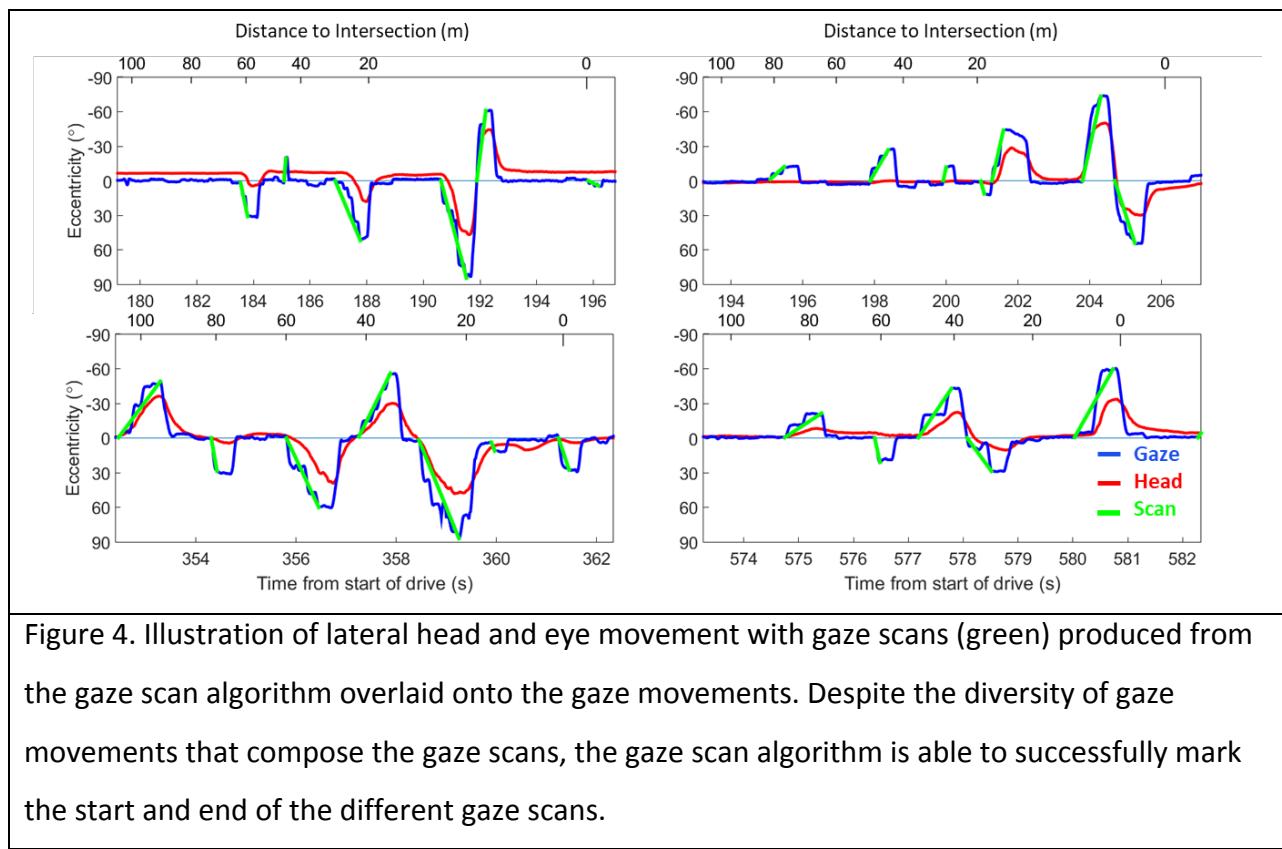
335 *Rule 1:* Both saccades must be on the same side of the straight ahead position, such that
 336 no saccades on the left side were merged with saccades on the right side, or vice versa. This
 337 rule prevented merging when two saccades were on opposite sides but satisfy the remaining
 338 rules.

339 *Rule 2:* Both saccades must be headed in the same direction (i.e. to the left, or to the
 340 right). This rule helped ensure that the end points of gaze scans were at the maximum
 341 eccentricity from the straight ahead position. Note that if two saccades qualified for merging
 342 but were separated by an intermediate saccade that did not satisfy this rule, the saccades may

343 still be merged assuming they satisfied *Rule 1* and *Rule 4* (see appendix section A.2 for example
 344 of how this is achieved).

345 *Rule 3:* The magnitude of the starting eccentricity of the later saccade must be greater
 346 than the magnitude of the starting eccentricity for the earlier saccade. The same must be true
 347 of the ending eccentricity as well. This rule helped ensure that each gaze scan included the
 348 maximum deviation from the straight ahead position and prevented unnecessary merging
 349 between likely distinct gaze scans.

350 *Rule 4:* The two saccades must be close in time to each other. The time that was
 351 selected, 0.4 s, is discussed in greater detail in section 3.4. If the difference in time between the
 352 end of the first saccade and the start of the second saccade exceeded this 0.4 s criterion, then
 353 the saccades were not merged. Given that gaze scans can occur sequentially on the same side
 354 (e.g. the multiple leftward scans on the top right in Figure 4), this rule prevents neighboring, yet
 355 separate, gaze scans from being merged together.



357 Merging was achieved by chronologically merging saccades until there were no more
 358 saccades that could be merged. This was achieved by repeating the merging procedure until
 359 there were two consecutive iterations with the same number of saccades. The remaining
 360 saccades (both those that were merged and not merged) were then treated as the final gaze
 361 scans. See appendix for a flowchart (section A.1) and written description (section A.2) of how
 362 the gaze scan algorithm steps through gaze data.

363 3.4 Optimizing the merging parameter

364 Rule 4 of the gaze scan algorithm determines how close in time two saccades need to be
 365 in order to be merged. We used the training set (Table 1) to optimize this parameter. The
 366 current parameter (i.e. 0.4 s) was selected by maximizing the product of the proportion of one-
 367 to-one gaze scan matches between the ground truth and the gaze scan algorithm and Cohen's
 368 Kappa (see section 3.6 for calculation) for each parameter value between 0.016 s to 0.750 in
 369 steps of 0.016 s (i.e., 1 sample at 60hz). A one-to-one match was defined as situations in which
 370 a single gaze scan from the ground truth was matched to a single gaze scan from the algorithm.
 371 Only for one-to-one matches could we evaluate the start and end markings of the gaze scan
 372 algorithm. Cases where more than one gaze scan was matched to a single gaze scan were
 373 labeled as one-to-many and many-to-one. One-to-many refers to situations where there were
 374 multiple algorithm gaze scans for a single ground truth gaze scan and many-to-one refers to
 375 situations where there were multiple ground truth gaze scans for a single algorithm gaze scan.
 376 As expected, increasing the time between saccades decreases the number of one-to-many
 377 errors and increases the number of many-to-one errors (Figure 5)

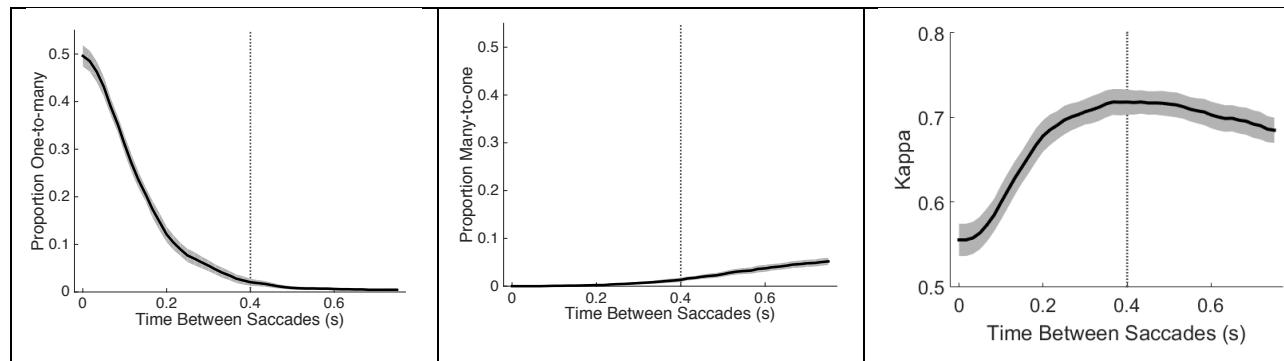


Figure 5. The effect that the value determining the maximum time between saccades (i.e. Rule 4) has on the proportion of one-to-many errors (left), on the proportion of many-to-one errors (middle), and on Cohen's Kappa (right; note: the graph is truncated at 0.5). The solid black line is the average for the 8 participants and the gray shading around the average represents the standard error. The vertical dotted line represents the value (0.4 s) that maximizes the product of 1 minus the proportion of one-to-many and many-to-one errors (i.e., proportion of one-to-one matches) and Cohen's Kappa.

378

379 3.5 Characterizing saccades and gaze scans generated by the gaze scan algorithm

380 Saccades and gaze scans generated by the gaze scan algorithm were characterized in
 381 terms of duration and magnitude. In addition, for gaze scans, the number of saccades per gaze
 382 scan was computed. The relationship between the duration and magnitude of saccades and
 383 gaze scans was quantified with Pearson correlations. Differences between the distributions of
 384 the durations and magnitudes of saccades and gaze scans were analyzed using two-sample
 385 Kolmogorov-Smirnov tests (given the non-normal distributions for gaze scan duration and
 386 magnitude). The relationship between the number of saccades per gaze scan and magnitude
 387 and duration was quantified with a series of Pearson correlations.

388 3.6 Quantifying performance of the gaze scan algorithm compared to the ground truth

389 To measure how well the gaze scan algorithm marked gaze scans, gaze scans from the
 390 algorithm were compared to the ground truth gaze scans from the testing set. We used a
 391 sample-by-sample Cohen's Kappa (K ; Cohen, 1960; Andersson, et al., 2017) to measure the
 392 reliability of the algorithm by comparing the relative observed agreement (P_o) and the
 393 hypothetical probability of chance agreement (P_e) of gaze data being marked as part of a gaze
 394 scan or not using the following formula:

$$395 \quad K = \frac{P_o - P_e}{1 - P_e}$$

396 Where $K = 1$ corresponds to perfect agreement and $K = 0$ corresponds to chance agreement.
397 Pearson correlations were used to estimate the relationship between the algorithm and ground
398 truth gaze scan durations and magnitudes. However, strong correlations do not necessarily
399 imply good agreement between two methods (in this case, gaze scan algorithm and ground
400 truth), especially if there is an offset in one method. Therefore, we used Bland-Altman methods
401 (Bland & Altman, 1986), which provide a way to investigate systematic differences between
402 two methods using the bias and variance (i.e., limits of agreement). These methods are more
403 sensitive than other methods (e.g., correlation, Cohen's Kappa) because the direction of the
404 bias can be ascertained and we can individually evaluate how well the algorithm is marking the
405 start and end time and eccentricity. We calculated both the bias and limits of agreement (LoA)
406 of the differences in duration and magnitude between the gaze scan algorithm and ground
407 truth. The significance of the bias was calculated using a sign-test, given that the differences in
408 duration and magnitude were not normally distributed in one-sample Kolmogorov-Smirnov
409 tests. LoAs were calculated by adding the median of the differences to the 2.5th and 97.5th
410 percentile. Effect sizes (r) for the sign-test were calculated by dividing the sign-test statistic (z)
411 by the square root of the sample size (Rosenthal, 1994). Bland-Altman methods were also used
412 for quantifying the differences between the start time, end time, start eccentricity, and end
413 eccentricity, in the same manner as for duration and magnitude.

414 When comparing the gaze scan algorithm to the ground truth, only those gaze scans
415 marked by the algorithm that could be paired with exactly one ground truth gaze scan (i.e.,
416 one-to-one matches) were analyzed, which corresponded to 92.5% of the marked gaze scans.
417 Gaze scans categorized as one-to-many (2.4%), many-to-one (2.6%), or had no corresponding
418 algorithm markings (2.5%) were not analyzed for the quality of their marking. However, it is
419 worth noting that the few one-to-many and many-to-one errors suggest that the gaze scan
420 algorithm successfully matched saccades according to the ground truth.

421 To evaluate the gaze scan algorithm, we compared the LoA between the gaze scan
422 algorithm and ground truth to the LoA between the three manual coders' manual markings of
423 the 'coders set' of data. The same methods to generate LoAs between the gaze scan algorithm

424 and ground truth were calculated for each coder compared to the other. Next, we averaged the
425 LoAs between the manual coders. This average is thus the difference we may expect between
426 manual coders, which provides a benchmark to determine whether the algorithm is performing
427 as well, worse, or the same as what we may expect for manual coders. See appendix (Appendix
428 A.4) for differences between manual coders.

429 To calculate the 95% confidence intervals around the LoAs, we utilized bootstrapped
430 resampling given the non-normality of the data. In 1000 iterations, we randomly selected, with
431 resampling, from the distribution until we had selected the same number of resamples as the
432 original distribution. Then, we calculated the LoAs for each iteration, thereby creating a
433 resampled distribution. The 95% confidence interval of the LoAs was defined by taking the 2.5%
434 and 97.5% percentile of the resampled distribution.

435

436 **4. Results**

437 4.1 Saccades and gaze scans generated by the gaze scan algorithm

438 The magnitude and duration of saccades [$r^2 = 0.63$, $p < 0.001$] and gaze scans [$r^2 = 0.43$,
439 $p < 0.001$] were found to be significantly correlated (Figure 6), similar to main sequence
440 relationships reported for eye saccades (Bahill, Clark, & Stark, 1975). As expected, the
441 distributions of the durations [$D = 0.63$, $p < 0.001$] and magnitudes [$D = 0.37$, $p < 0.001$] were
442 significantly different between saccades and gaze scans. Saccades had smaller durations with
443 less dispersion [median = 0.07, IQR = 0.04 s to 0.083] than gaze scans (median = 0.24, IQR = 0.09
444 s to 0.45 s). The same was also true of magnitudes [saccades: median = 4.3, IQR = 2.1° to 10.0°;
445 gaze scans: median = 12.7, IQR = 6.2° to 36.2°]. Longer duration and larger magnitude gaze
446 scans compared to saccades was expected given that gaze scans could be composed of multiple
447 saccades.

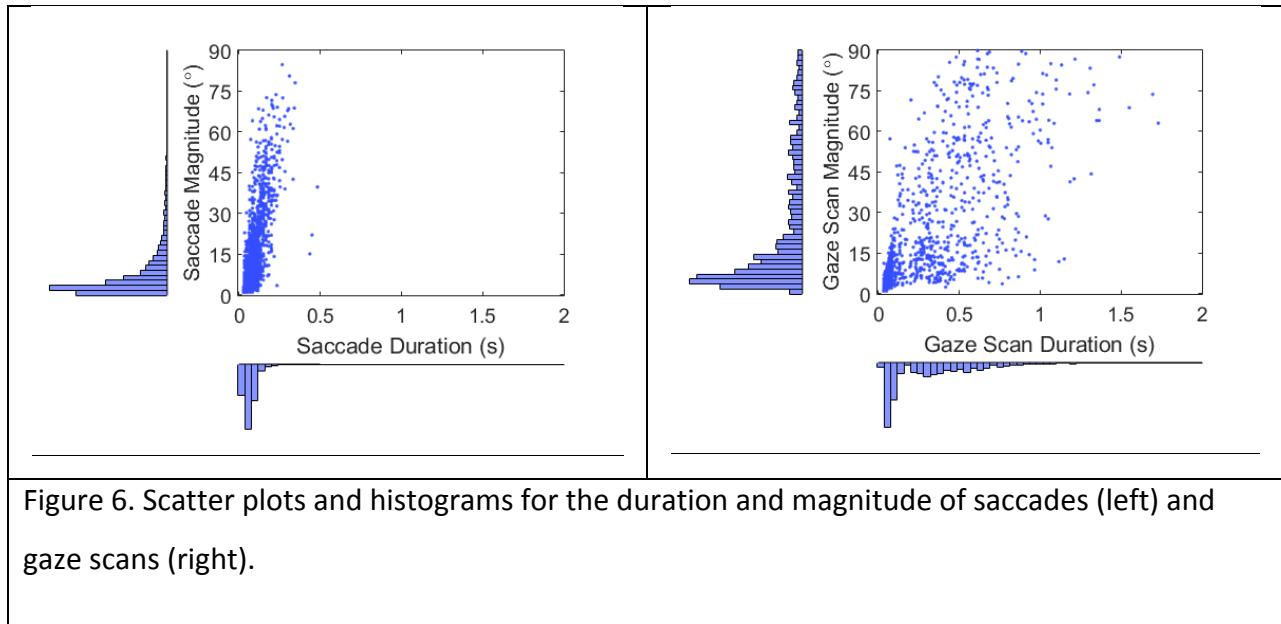


Figure 6. Scatter plots and histograms for the duration and magnitude of saccades (left) and gaze scans (right).

448

449 Approximately 55.2% of gaze scans were composed of more than one saccade (Figure 7
 450 left). The duration [$r^2 = 0.84, p < 0.001$] and magnitude [$r^2 = 0.44, p < 0.001$] of gaze scans were
 451 significantly positively correlated with the number of saccades per gaze scan (Figure 7 center
 452 and right, respectively). This was expected given that individuals typically don't make many eye
 453 saccades greater than 15° (Bahill, Adler, & Stark, 1975) and larger gaze scans would, therefore,
 454 require more saccades. Finding a majority of the gaze scans are composed of multiple saccades
 455 and that the number of saccades affects both the magnitude and duration of gaze scans
 456 supports the usefulness of the gaze scan algorithm when merging gaze scans together.

457

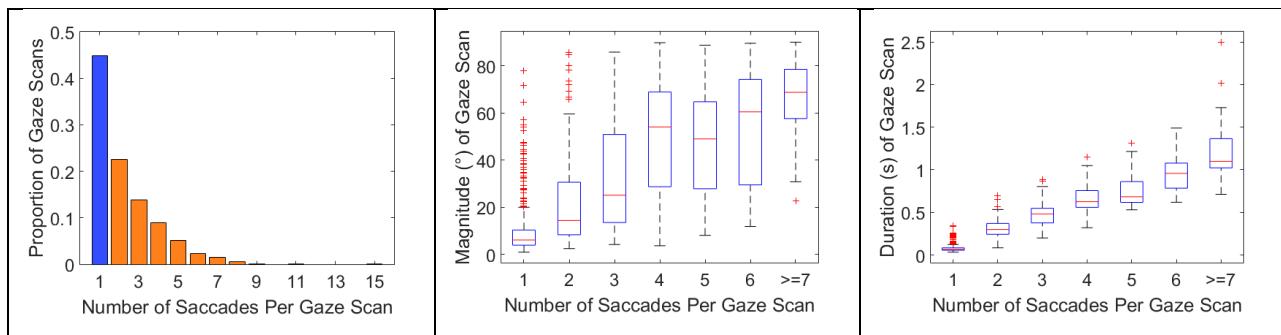


Figure 7. Proportion of gaze scans with specific numbers of saccades (left) with blue representing gaze scans composed of a single saccade and orange representing those

composed of more than one. The magnitude of gaze scans as a function of the number of saccades within each gaze scan (middle). The duration of gaze scans as a function of the number of saccades (right)

458

459 4.2 Comparing gaze scans between the gaze scan algorithm and ground truth

460 Gaze scan duration [$r^2 = 0.61$, $p < 0.001$] and gaze scan magnitude [$r^2 = 0.995$, $p < 0.001$]
 461 were significantly positively correlated (Figure 8 left) between the gaze scan algorithm and
 462 ground truth, with the relationship being stronger for magnitude than duration [$z = 50.7$, $p <$
 463 0.001]. The sample-to-sample Cohen's kappa for all gaze scans between the algorithm and
 464 ground truth was 0.62, which suggests good agreement (Cohen, 1960) and is similar to the
 465 sample-to-sample kappa between expert coders in this study (see Table A.1) and found for
 466 other saccade detection algorithms (Andersson, et al., 2017; 60hz data in Zemblys, et al., 2018)

467 The differences in duration [$p < 0.001$] and magnitude [$p < 0.001$] were found to be
 468 significantly different from a normal distribution, which was likely due to the distributions being
 469 highly leptokurtic [kurtosis for durations = 24.6, magnitudes = 56.4, standard error of kurtosis =
 470 0.16]. When evaluating agreement between the gaze scan algorithm and ground truth with the
 471 Bland-Altman methods (Figure 8 right), the duration [median = -0.01 s, $z = 6.0$, $p < 0.001$] was
 472 significantly biased towards the ground truth, albeit with a small effect size [$r = 0.19$] and a bias
 473 that is smaller than what can be measured with our system (i.e., our sampling rate was 60 hz).
 474 The magnitude was not significantly biased [median = 0.02°, $z = 0.5$, $p = 0.63$] towards either
 475 the gaze scan algorithm or ground truth.

476 The comparisons of the limits of agreement (LoA) between the algorithm and ground
 477 truth are summarized in Table 2 and described further below.

	Algorithm vs. Ground Truth	Inter-coder
Magnitude (°)	3.82 (3.36 to 4.45)	3.42 (2.95 to 4.05)

Duration (s)	0.41 (0.37 to 0.46)	0.29 (0.24 to 0.34)
Start gaze scan eccentricity (°)	2.27 (1.88 to 2.63)	1.65 (1.37 to 1.95)
End gaze scan eccentricity (°)	2.42 (2.15 to 3.33)	2.74 (2.3 to 3.35)
Start gaze scan time (s)	0.28 (0.24 to 0.32)	0.17 (0.12 to 0.21)
End gaze scan time (s)	0.28 (0.25 to 0.33)	0.22 (0.19 to 0.26)

Table 2. Average limits of agreement (LoA) between the gaze scan algorithm and ground truth and between the coders. 95% confidence intervals are displayed inside the parentheses.

478

479 The LoA for magnitude between the gaze scan algorithm and ground truth were within
 480 the average confidence interval of the LoA between manual coders (Table 2), which suggests
 481 that the level of agreement between the algorithm and ground truth was similar to that found
 482 between expert coders. However, this was not the case for the LoAs for duration, given that
 483 confidence intervals between the algorithm and ground truth and manual coders did not
 484 overlap (Table 2). Despite the lack of an overlap in LoAs for duration, approximately 90.5% of
 485 differences between the gaze scan algorithm and ground truth were within the lower and
 486 upper confidence bounds between the manual coders, suggesting that the wider LoA between
 487 the algorithm and ground truth was driven by a few outliers in durations.

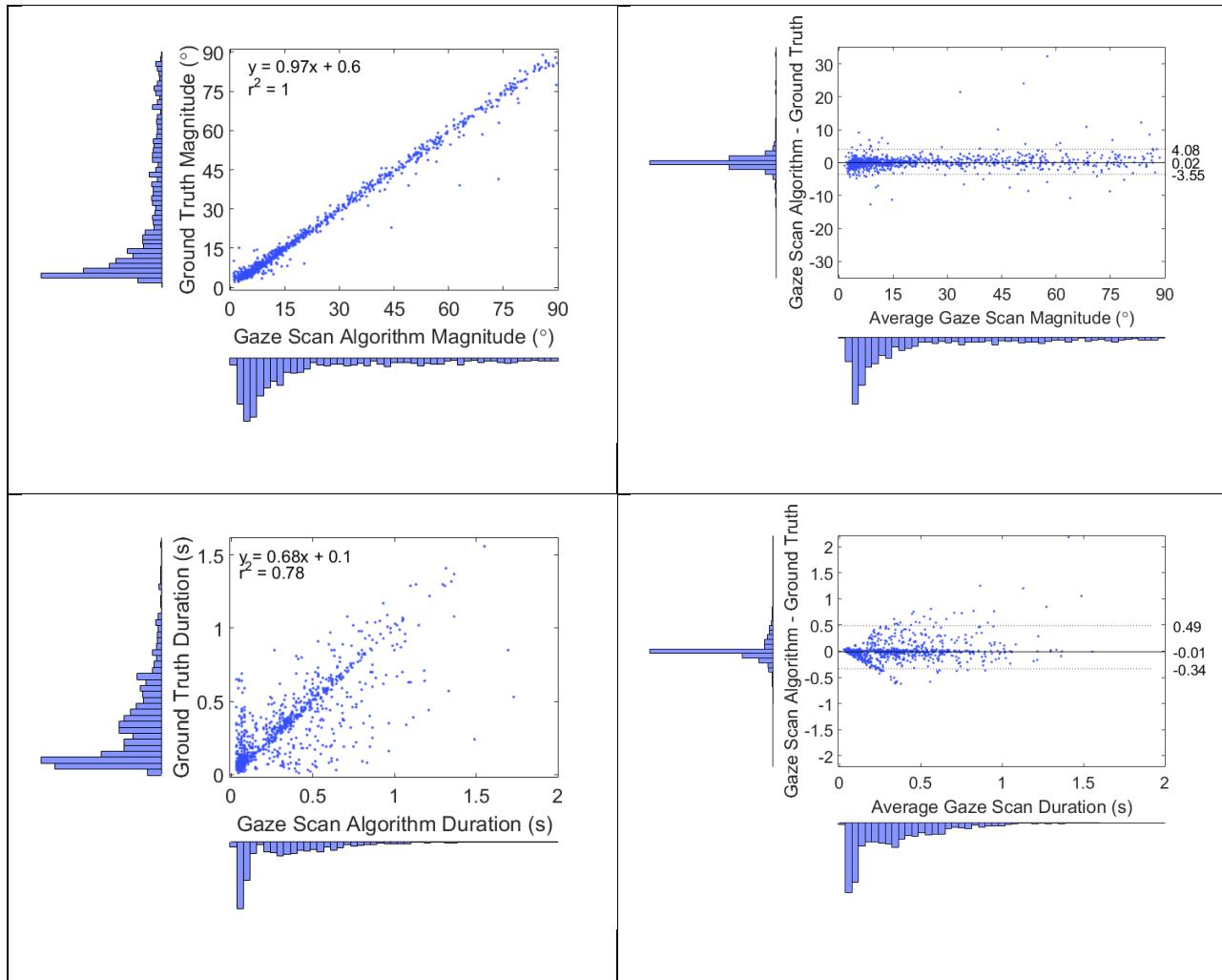


Figure 8. Scatterplots and histograms showing the relationship between the gaze scan algorithm and ground truth magnitudes (top left) and durations (bottom left). Bland-Altman plots showing the difference between the algorithm and ground truth magnitudes (top right) and duration (bottom right). The dotted horizontal lines represent the limits of agreement (LoA) and the numbers correspond to those limits with the median between the two LoAs.

488

489 As was the case with duration and magnitude, the error distributions for start time [$p <$
 490 0.001], end time [$p < 0.001$], start eccentricity [$p < 0.001$], and end eccentricity [$p < 0.001$]
 491 between gaze scans from the algorithm and ground truth were found to be significantly
 492 different from a normal distribution. The non-normality was likely related to the distributions

493 being highly leptokurtic (kurtosis: start time = 27.1, end time = 53.4, start eccentricity = 22.2,
 494 end eccentricity = 96.0, standard error of kurtosis = 0.16). Bland-Altman plots for the
 495 differences of start time, end time, start eccentricity and end eccentricity between the
 496 algorithm and ground truth are displayed in Figure 9. The difference in end time was
 497 significantly biased towards the ground truth [median = -0.01 s, $z = 6.6$, $p < 0.001$], albeit with a
 498 small effect size [$r = 0.21$]. However, the difference in start time [median = 0.0 s, $z = 0.6$, $p =$
 499 0.54], start eccentricity [median = 0.0°, $z = 1.5$, $p = 0.13$], and end eccentricity [median = 0.3°, z
 500 = 1.4, $p = 0.15$] were not significantly biased.

501

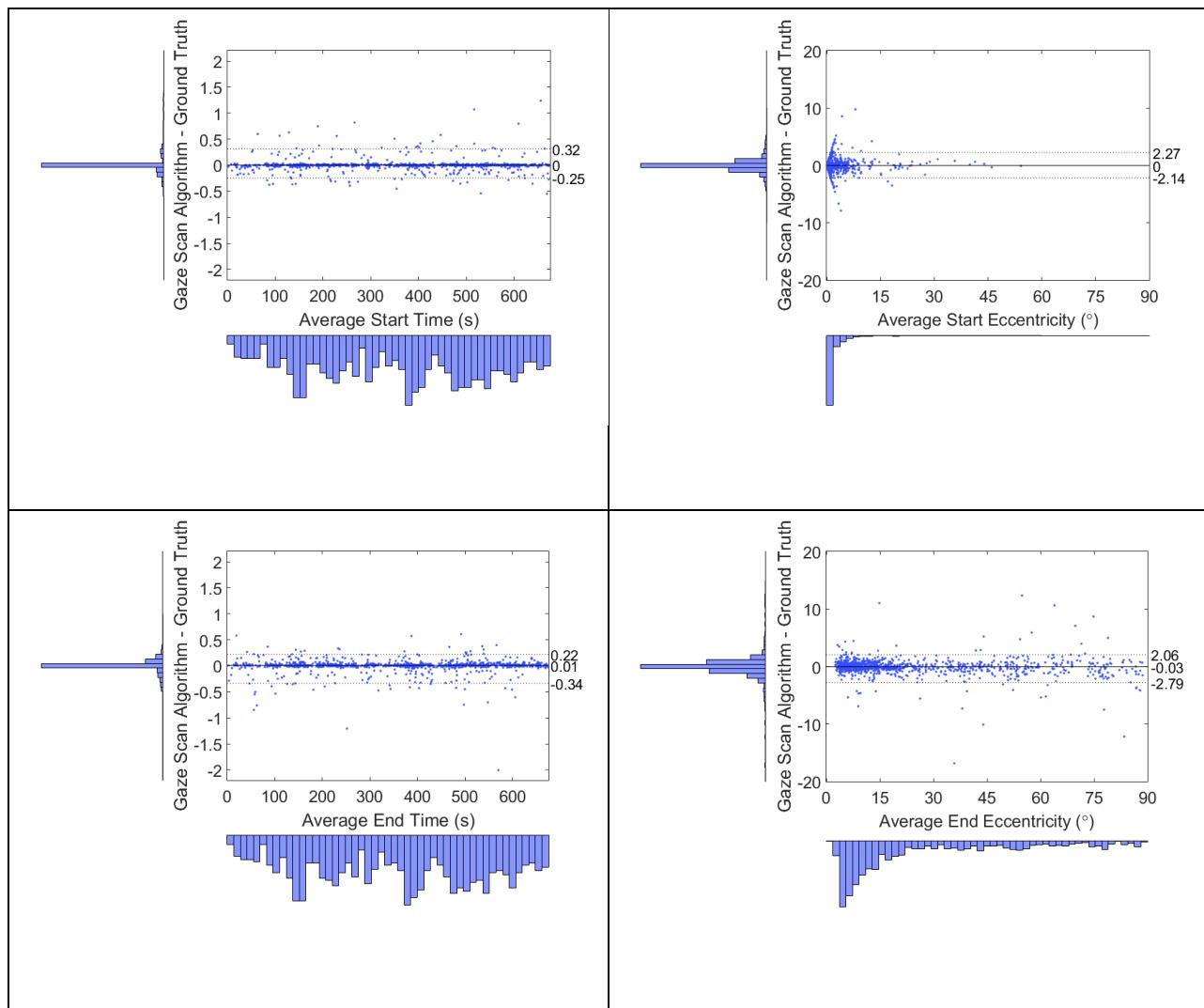


Figure 9. Differences between the gaze scan algorithm and ground truth for each matched gaze scan's start time (top left), end time (bottom left), starting eccentricity (top right), and ending eccentricity (bottom right). The dotted horizontal lines represent the limits of agreement (LoA) and the numbers correspond to those limits with the median between the two LoAs.

502

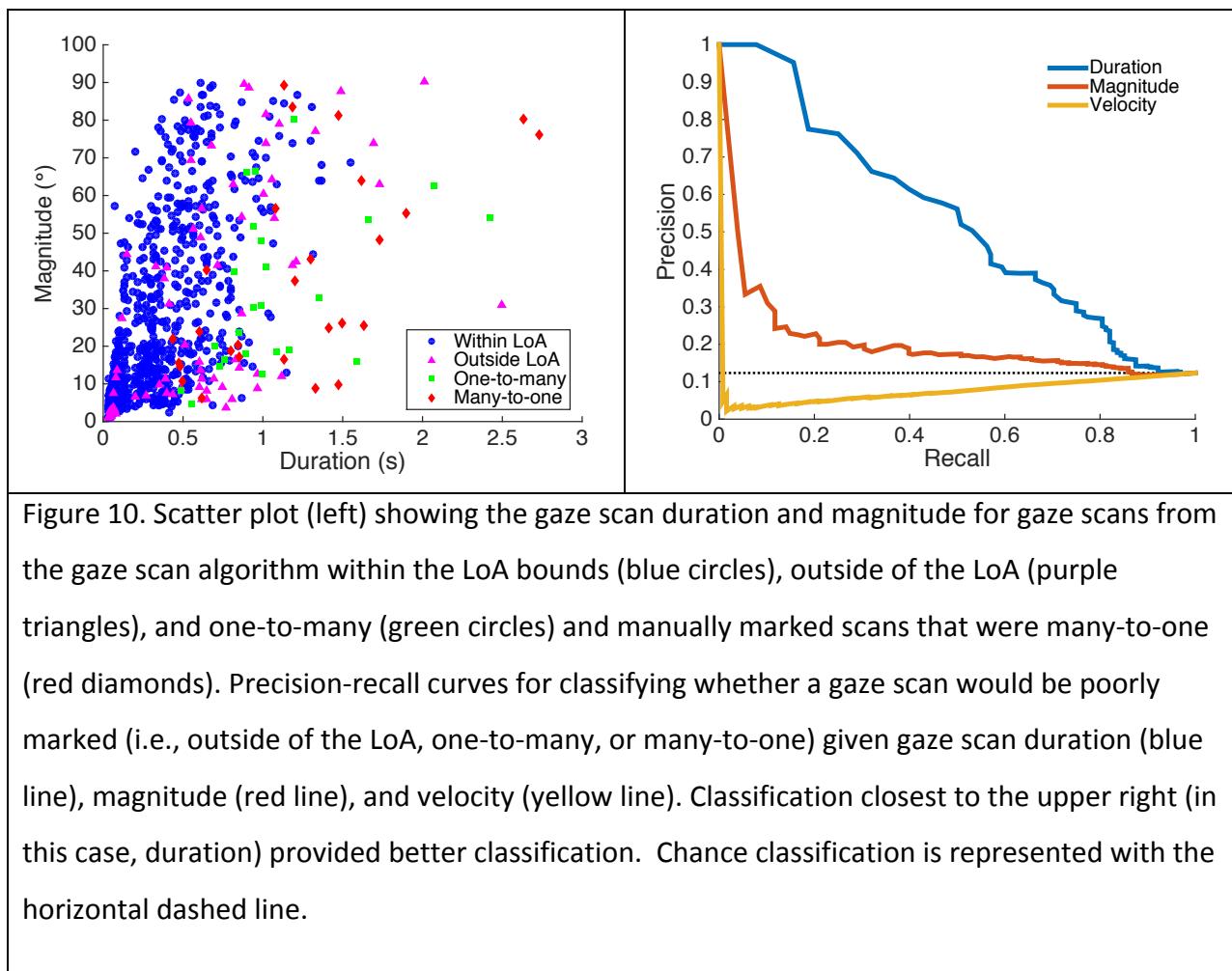
503 The LoAs between the gaze scan algorithm and ground truth for end eccentricity overlap
504 with the average confidence intervals of the LoAs between the manual coders for end
505 eccentricity (Table 2), suggesting agreement between algorithm and manual coders regarding
506 where the gaze scan ends in eccentricity. There was some overlap for start eccentricity and end
507 time, but no overlap for start time (Table 2). As was the case with duration, 92.4% of the
508 differences between the algorithm start times were within the lower and upper confidence
509 bounds between the manual coders, suggesting a few outliers may have been driving the worse
510 agreement between gaze scan algorithm and ground truth.

511

512 4.3 Addressing gaze scans poorly marked by the algorithm

513 As is the case in any event detection algorithm, the goal is to accelerate processing of
514 gaze data without sacrificing accuracy. As identified here, the gaze scan algorithm produced
515 one-to-many errors (2.4%) and many-to-one errors (2.6%) when compared to the ground truth.
516 These gaze scans, and gaze scans with a duration or magnitude that were outside the ground
517 truth LoA (approximately 7.8%, Figure 10), could then be manually inspected and corrected
518 where necessary. However, without manual marking, it would be difficult to know in advance
519 which gaze scans are poorly marked. We utilized precision-recall curves to evaluate whether
520 gaze scan duration, magnitude, or velocity may be predictors of poor fitting. Precision-recall
521 curves are similar to receiver operator characteristic (ROC) curves, except that precision-recall
522 curves are more appropriate for imbalanced datasets (Saito & Rehmsmeier, 2015). Unlike ROC
523 curves, better classification corresponds to recall and precision closest to 1 (i.e., towards the
524 upper right). Area under the curve (AUC), which summarizes classification performance, was

estimated using the trapezoidal rule. AUCs for classifying poorly fit gaze scans were 0.53, 0.21, and 0.08 for gaze scan duration, magnitude, and velocity, respectively. For gaze scan duration, the threshold that best separated true and false positive rates was 0.6 s, which suggests that that threshold may be useful in indicating whether a gaze scan may be poorly marked. Specifically, this threshold may be most useful in capturing one-to-many and many-to-one errors (i.e., 92% and 82% were above 0.6 s, respectively) and less useful for gaze scans outside of the LoA (48.7%).



532

533 5. Discussion

534 We developed an algorithm to automatically detect gaze (head combined with eye
535 eccentricity) scans by marking the start and end of each scan called the gaze scan algorithm.

536 We compared performance of the algorithm to a ground-truth dataset of manually-marked
537 scans. In addition, we compared the differences between the gaze scan algorithm and manually
538 marked scans to differences found between expert coders to better understand what may be
539 considered adequate markings by the algorithm.

540 The algorithm's primary function is to merge saccades into gaze scans. To determine if
541 this was necessary, we calculated the number of saccades per gaze scan to determine how
542 frequently gaze scans were composed of multiple saccades. Approximately 55.2% of the
543 matched gaze scans in the testing set were composed of multiple saccades, suggesting that the
544 algorithm was necessary in marking the full extent of the gaze scans. For the testing set, less
545 than 2.4% of the ground truth gaze scans were one-to-many by the gaze scan algorithm
546 compared to 49.4% in a version of the algorithm without any merging. These results suggest
547 that the algorithm successfully merged multiple saccades into gaze scans.

548 Overall, the gaze scan algorithm and ground truth produced qualitatively and
549 quantitatively similar gaze scans. In the testing set, 95% of the gaze scans produced by the
550 algorithm were matched to a gaze scan from the ground truth data set, suggesting the
551 algorithm successfully marked gaze scans. When considering the magnitude and duration of the
552 gaze scans, there was good agreement according to Cohen's Kappa and significant correlations
553 between the algorithm and ground truth gaze scans for both the magnitude and duration,
554 albeit with a stronger correlation for magnitude than duration. In addition, we assessed the
555 agreement between the gaze scan durations and magnitudes between the algorithm and
556 ground truth using limits of agreement (LoA) from Bland-Altman methods. The agreement
557 between the gaze scan algorithm and ground truth for gaze scan magnitude was similar to the
558 agreement between the expert coders, suggesting that the algorithm is sufficiently marking the
559 magnitude of the gaze scan. Furthermore, similar results were found for both the start and end
560 eccentricity and end timing. However, there was less agreement for gaze scan duration
561 between the gaze scan algorithm and ground truth compared to the expert coders. When
562 examining the agreement between the gaze scan algorithm and ground truth for start and end
563 times, there was less agreement for start times than end times, which may explain the

564 variability for durations produced by the gaze scan algorithm. However, even though there was
565 less agreement between the gaze scan algorithm and ground truth for the timing of gaze scans,
566 more than 90% of the gaze scans were still within the agreement range of the manual coders.
567 Thus, the gaze scan algorithm and ground truth tended to agree about as well as expert manual
568 coders tend to agree. Gaze scan duration is one metric (section 4.3) that may be useful in
569 identifying gaze scans that may be poorly marked by the gaze scan algorithm and need to be
570 corrected with manual marking. In our dataset, gaze duration exceeding 0.6 s seemed to be a
571 reasonable threshold, though this value may change based on the driving scenario.

572 The current implementation of the gaze scan algorithm focused on quantifying gaze
573 scanning on approach to intersections, but could also be applied to scanning in other driving
574 scenarios. It is applicable to different driving environments that may have different types of
575 scanning, such as driving on the highway versus driving in the city. The algorithm complements
576 existing research measuring for how long or how frequently individuals look at different
577 sections of the road (Yamani, Samuel, Gerardino, & Fisher, 2016), hazards (Crundall et al., 2012)
578 or at in-vehicle displays (Donmez, Boyle, & Lee, 2009) by providing a way to quantify how
579 individuals moved their eyes to reach that area of interest. In addition, the algorithm could be
580 used to determine the magnitude of gaze scans when walking; for example, determining when
581 it is safe to cross a street requires large gaze scans to the left and right (e.g. Whitebread &
582 Neilson, 2000; Hassan, Geruschat, & Turano, 2005). In applied settings, the algorithm could be
583 used to quantify an individual's scanning behaviors (how far and how frequently they scanned)
584 to monitor progress during scanning training as part of a rehabilitation program for drivers who
585 exhibited scanning deficits, such as individuals with visual field loss (Bowers et al., 2014) or
586 older persons with normal vision (Romoser & Fisher, 2009).

587 One potential limitation of the gaze scan algorithm is that detecting saccadic gaze
588 movements using velocity thresholds at low sampling rates (i.e. less than 250 Hz) results in
589 imprecise markings (Mack, Belfanti, & Schwarz, 2017). Therefore, the accuracy and
590 optimization of the algorithm may have been impacted by imprecise markings of saccades
591 because the gaze data used was collected at 60 Hz. While the 60 Hz sampling rate might have

592 influenced the accuracies described here, the algorithm is not dependent upon the sampling
593 rate and can be considered modular. That is, the merging portion of the algorithm (i.e. Stage 2
594 described in section 3.3.3) could be applied to saccades detected from a different algorithm
595 using a different sampling rate from the methodology used in Stage 1 described in this paper.

596 While the current configuration of the gaze scan algorithm sufficiently marked gaze
597 scans compared to the ground truth scans, it is possible that there may be subgroups of
598 participants wherein a different configuration of the algorithm would provide a better fit of
599 data. For example, age impacts how an individual scans when driving on-road (Bao & Boyle,
600 2009b) and in the driving simulator (Romoser, Pollatsek, Fisher, & Williams, 2013; Savage et al.,
601 *Revise and Resubmit*) and this could mean that age may impact the parameter value that
602 determines how close in time two saccades need to be to be merged. With the current data set,
603 there is not enough data to determine whether this should be the case or the case for other
604 potential subgroups (e.g. gender, driving experience).

605 **6. Conclusion**

606 We describe an algorithm that automatically marks the beginning and end of lateral gaze scans,
607 which allows for the quantification of the duration, magnitude, and composition of those scans,
608 called the gaze scan algorithm. The algorithm produces gaze scans that are quantitatively
609 similar in duration and magnitude to manually marked ground truth gaze scans with differences
610 from the ground truth within the level of agreement that may be expected between expert
611 manual coders. Therefore, the algorithm may be used in lieu of manual marking of gaze data,
612 significantly accelerating the time consuming marking of gaze movement data in driving
613 simulator studies. The algorithm complements existing driving simulator research investigating
614 the relationships between gaze movements and driving behavior and could be implemented in
615 other situations outside of the driving simulator (e.g. walking) that involve multiple gaze
616 movements headed in the same direction.

617

618 **Acknowledgements**

619 The research was supported in part by NIH grant R01-EY025677, NIH Core Grant
620 P30EY003790, and S10 – RR028122

621

622 **Author Contributions**

623 All authors contributed to the manuscript. B.G. and A.R.B. originally conceived of the
624 gaze scan algorithm. G.S., B.G., A.A. and L.Z. contributed to the programming of the algorithm.
625 G.S., S.W.S., and L.Z. contributed to initial parameter optimization. G.S. and A.A. developed the
626 methodology for optimizing the merging parameters.

627

628 **Open Access Statement**

629 Code for the gaze scan algorithm and manual marking can be downloaded from
630 <https://osf.io/p6jqn/>. Furthermore, the data from the training, testing, and coders sets can also
631 be downloaded from that same location.

632

633 **References**

- 634 Alberti, C. F., Goldstein, R. B., Peli, E., & Bowers, A. R. (2017). Driving with hemianopia V: do
635 individuals with hemianopia spontaneously adapt their gaze scanning to differing hazard
636 detection demands?. *Translational vision science & technology*, 6(5), 11-11.
- 637 Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., & Nyström, M. (2017). One algorithm to
638 rule them all? An evaluation and discussion of ten eye movement event-detection
639 algorithms. *Behavior research methods*, 49(2), 616-637.
- 640 Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have
641 magnitudes of 15 degrees or less. *Investigative Ophthalmology & Visual Science*, 14(6), 468-469.
- 642 Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye
643 movements. *Mathematical Biosciences*, 24(3-4), 191-204.

- 644 Bahnemann, M., Hamel, J., De Beukelaer, S., Ohl, S., Kehrer, S., Audebert, H., ... & Brandt, S. A.
645 (2015). Compensatory eye and head movements of patients with homonymous hemianopia in
646 the naturalistic setting of a driving simulation. *Journal of neurology*, 262(2), 316-325.
- 647 Bao, S., & Boyle, L. N. (2009a). Driver safety programs: The influence on the road performance
648 of older drivers. *Transportation Research Record: Journal of the Transportation Research Board*,
649 (2096), 76-80.
- 650 Bao, S., & Boyle, L. N. (2009b). Age-related differences in visual scanning at median-divided
651 highway intersections in rural areas. *Accident Analysis & Prevention*, 41(1), 146-152.
- 652 Barnes, G. R. (1979). Vestibulo-ocular function during co-ordinated head and eye movements to
653 acquire visual targets. *The Journal of Physiology*, 287(1), 127-147.
- 654 Beintema, J. A., van Loon, E. M., & van den Berg, A. V. (2005). Manipulating saccadic decision-
655 rate distributions in visual search. *Journal of vision*, 5(3), 1-1.
- 656 Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two
657 methods of clinical measurement. *The lancet*, 327(8476), 307-310.
- 658 Bowers, A. R., Ananyev, E., Mandel, A. J., Goldstein, R. B., & Peli, E. (2014). Driving with
659 hemianopia: IV. Head scanning and detection at intersections in a simulator. *Investigative*
660 *ophthalmology & visual science*, 55(3), 1540-1548..
- 661 Bowers, A. R., Bronstad, M. P., Spano, L. P., Goldstein, R. BB., & Peli, E. (2019). The effects of
662 age and central field loss on head scanning and detection at intersections. *Translational Vision*
663 *Science & Technology*, 8(5), 14-14.
- 664 Braitman, K. A., Kirley, B. B., McCartt, A. T., & Chaudhary, N. K. (2008). Crashes of novice
665 teenage drivers: Characteristics and contributing factors. *Journal of safety research*, 39(1), 47-
666 54.
- 667 Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological*
668 *measurement*, 20(1), 37-46.
- 669 Crundall, D., Chapman, P., Trawley, S., Collins, L., Van Loon, E., Andrews, B., & Underwood, G.
670 (2012). Some hazards are more attractive than others: Drivers of varying experience respond
671 differently to different types of hazard. *Accident Analysis & Prevention*, 45, 600-609.
- 672 Donmez, B., Boyle, L. N., & Lee, J. D. (2009). Differences in off-road glances: effects on young
673 drivers' performance. *Journal of transportation engineering*, 136(5), 403-409.

- 674 Guitton, D., & Volle, M. (1987). Gaze control in humans: eye-head coordination during orienting
675 movements to targets within and beyond the oculomotor range. *Journal of*
676 *neurophysiology*, 58(3), 427-459.
- 677 Hakamies-Blomqvist, L. E. (1993). Fatal accidents of older drivers. *Accident Analysis &*
678 *Prevention*, 25(1), 19-27.
- 679 Hamel, J., De Beukeleir, S., Kraft, A., Ohl, S., Audebert, H. J., & Brandt, S. A. (2013). Age-related
680 changes in visual exploratory behavior in a natural scene setting. *Frontiers in psychology*, 4, 339.
- 681 Hassan, S. E., Geruschat, D. R., & Turano, K. A. (2005). Head movements while crossing streets:
682 effect of vision impairment. *Optometry and vision science*, 82(1), 18-26.
- 683 Keskinen, E., Ota, H., & Katila, A. (1998). Older drivers fail in intersections: Speed discrepancies
684 between older and younger male drivers. *Accident Analysis & Prevention*, 30(3), 323-330.
- 685 Mack, D. J., Belfanti, S., & Schwarz, U. (2017). The effect of sampling rate and lowpass filters on
686 saccades—a modeling approach. *Behavior research methods*, 49(6), 2146-2162.
- 687 McKnight, A. J., & McKnight, A. S. (2003). Young novice drivers: careless or clueless?. *Accident*
688 *Analysis & Prevention*, 35(6), 921-925.
- 689 Munn, S. M., Stefano, L., & Pelz, J. B. (2008). Fixation-identification in dynamic scenes:
690 Comparing an automated algorithm to manual coding. In *Proceedings of the 5th symposium on*
691 *Applied perception in graphics and visualization* (pp. 33-42). ACM.
- 692 Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade
693 detection in eyetracking data. *Behavior research methods*, 42(1), 188-204.
- 694 Reimer, B. (2009). Impact of cognitive task complexity on drivers' visual
695 tunneling. *Transportation Research Record*, 2138(1), 13-19.
- 696 Romoser, M. R., & Fisher, D. L. (2009). The effect of active versus passive training strategies on
697 improving older drivers' scanning in intersections. *Human factors*, 51(5), 652-668.
- 698 Romoser, M. R., Fisher, D. L., Mourant, R., Wachtel, J., & Sizov, K. (2005). The use of a driving
699 simulator to assess senior driver performance: Increasing situational awareness through post-
700 drive one-on-one advisement. In: *Proceedings of the 8th International Driving Symposium on*
701 *Human Factors in Driver Assessment, Training and Vehicle Design*. Iowa City, IA: University of
702 Iowa, 239-245
- 703 Romoser, M. R., Pollatsek, A., Fisher, D. L., & Williams, C. C. (2013). Comparing the glance
704 patterns of older versus younger experienced drivers: Scanning for hazards while approaching

- 705 and entering the intersection. *Transportation research part F: traffic psychology and*
706 *behaviour*, 16, 104-116.
- 707 Rosenthal, R., Cooper, H., & Hedges, L. (1994). Parametric measures of effect size. *The*
708 *handbook of research synthesis*, 621(2), 231-244.
- 709 Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking
710 protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications* (pp.
711 71-78). ACM.
- 712 Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC
713 plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3).
- 714
- 715 Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least
716 squares procedures. *Analytical chemistry*, 36(8), 1627-1639.
- 717 Savage, S. W., Zhang, L., Pepo, D., Sheldon, S. S., Spano, L. P., & Bowers, A. R. (2017). The Effects
718 of Guidance Method on Detection and Scanning at Intersections—A Pilot Study. In *Proceedings*
719 *of the International Driving Symposium on Human Factors in Driver Assessment, Training, and*
720 *Vehicle Design*. Manchester Village, VT: University of Iowa, 340-346
- 721 Savage, S. W., Zhang, L., Swan, G., & Bowers, A. R. (Revise & resubmit). The effects of age on
722 the relative contributions of the head and the eyes to scanning behaviors at intersections.
723 *Transportation Research Part F: Traffic Psychology and Behaviour*
- 724 Smeets, J. B., & Hooge, I. T. (2003). The nature of variability in saccades. *Journal of*
725 *neurophysiology*.
- 726 Sodhi, M., Reimer, B., & Llamazares, I. (2002). Glance analysis of driver eye movements to
727 evaluate distraction. *Behavior Research Methods, Instruments, & Computers*, 34(4), 529-538.
- 728 Whitebread, D., & Neilson, K. (2000). The contribution of visual search strategies to the
729 development of pedestrian skills by 4-11 year-old children. *British Journal of Educational*
730 *Psychology*, 70(4), 539-557.
- 731 Yamani, Y., Samuel, S., Roman Gerardino, L., & Fisher, D. L. (2016). Extending analysis of older
732 drivers' scanning patterns at intersections. *Transportation Research Record: Journal of the*
733 *Transportation Research Board*, (2602), 10-15.
- 734 Zemblys, R., Niehorster, D. C., Komogortsev, O., & Holmqvist, K. (2018). Using machine learning
735 to detect events in eye-tracking data. *Behavior research methods*, 50(1), 160-181.
- 736

737 Appendix

A.1. Flow chart for processing gaze data using gaze scan algorithm

739

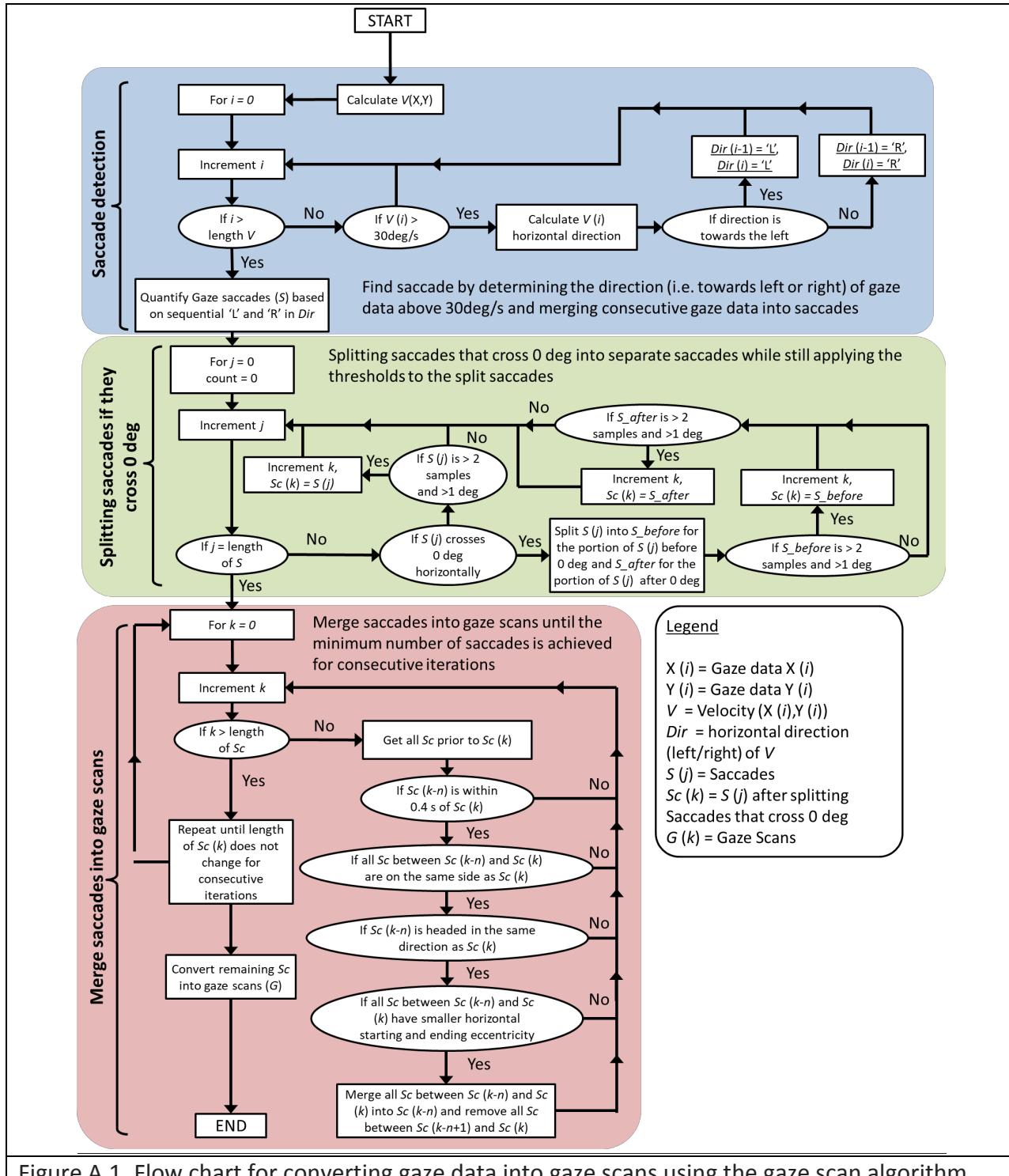


Figure A.1. Flow chart for converting gaze data into gaze scans using the gaze scan algorithm.

Here, $V(X,Y)$ refers to gaze velocity in the horizontal and vertical directions. Saccade is simplified to S, saccades after correcting for saccades that cross zero is simplified to Sc, and gaze scans are simplified to G.

740

741 A.2. Verbal description of the gaze scan algorithm

742 The procedure for merging saccades is described in a simplified format below. Figure A.2
 743 shows gaze data from Figures 3 and 5 (lower right plot) from the manuscript, to illustrate how
 744 the gaze scan algorithm merges saccades (Figure A.2, left) into gaze scans (Figure A.2, right). For
 745 the sake of simplicity, saccade is represented by S and the number next to S represents which
 746 saccade.

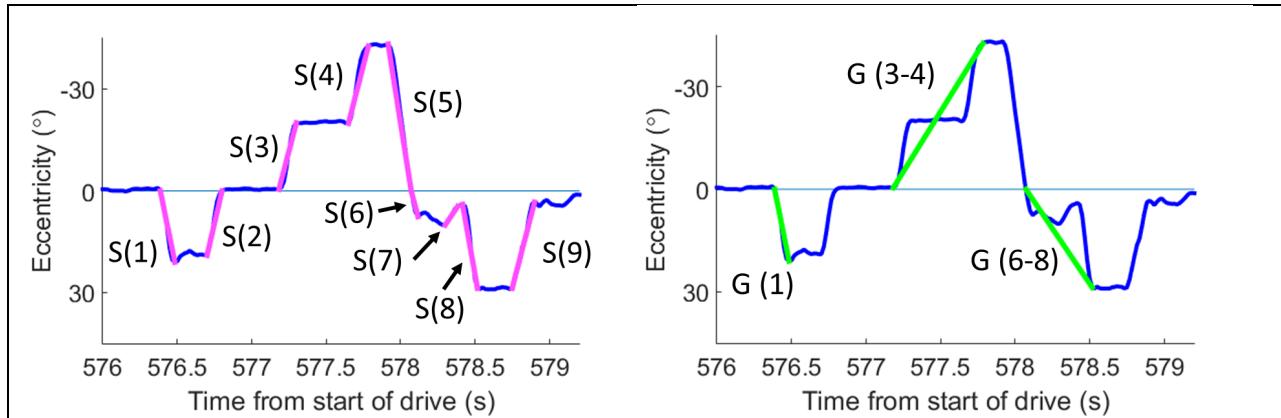


Figure A.2. Zoomed in data from Figures 3 and 5 (lower right plot) illustrating how saccades (on the left) are merged into gaze scans (on the right). On the left, S (N) represents the Nth saccade (S). On the right, G (N) represents the saccades that compose a gaze scan (G). Note that for simplicity, only large saccades are being shown.

747

748 Merging begins with S (2). Given that S (1) is headed in a different direction (i.e. towards
 749 the right) than S (2), S (2) is not merged with S (1) based on Rule 2. Next, we examine S (3).
 750 Given that S (1) and S (2) are on the opposite side (i.e. on the right) to S (3) (i.e. on the left), S
 751 (3) is not merged with either S (1) or S (2) based on Rule 1. Next, we examine S (4). Like S (3), S
 752 (4) cannot be merged with S (1) and S (2) because they are on opposite sides. However, the
 753 relationship between S (3) and S (4) satisfies all of the rules. Therefore, S (3) and S (4) are

754 merged into a single S (3-4), which has the start of S (3) and the end of S (4). Next, we examine
755 S (5). Given that S (1) and S (2) are on the opposite side to S (5) and that S (3-4) is headed in a
756 different direction than S (5), S (5) is not merged. Next, we examine S (6). Given that S (1) and S
757 (2) are separated from S (6) by more than 0.4 s (i.e. Rule 4) and that S (3-4) and S (5) are on the
758 opposite side to S (6), S (6) is not merged. Next, we examine S (7). Given that S (1) and S (2) are
759 too far back in time, S (3-4) and S (5) are the opposite side to S (7), and S (6) is headed in a
760 different direction, S (7) is not merged. Next, we examine S (8). Given that S (1) and S (2) are
761 too far back in time and S (3-4) and S (5) are on the opposite side, S (8) cannot be merged with
762 those saccades. However, S (6) is within the merging time set by Rule 4 and can be merged with
763 S (8). Furthermore, S (7), while headed in a different direction than S (8), is merged with S (6)
764 and S (8) because it is on the same side (i.e. Rule 1) and is sandwiched between S (6) and S (8),
765 thereby resulting in S (6-8) with the start of S (6) and end of S (8). Next, we examine S (9). Given
766 that S (1) and S (2) are too far back in time, S (3-4) and S (5) are on the opposite side, and S (6-8)
767 is headed in a different direction, S (9) is not merged. Now that all of the saccades have been
768 merged, the remaining saccades are treated as the final gaze scans. When only considering the
769 gaze scans headed away from 0°, S (1) becomes the first gaze scan G (1), S (3-4) becomes G (3-
770 4), and S (6-8) becomes G (6-8); the numbers in parentheses still refer to the original numbering
771 of the saccades.

772

773 A.3. Saccades markings

774 In Figure A.3, the saccade markings are displayed on the same data as displayed in
775 Figure 1 and Figure 4 in the main paper.

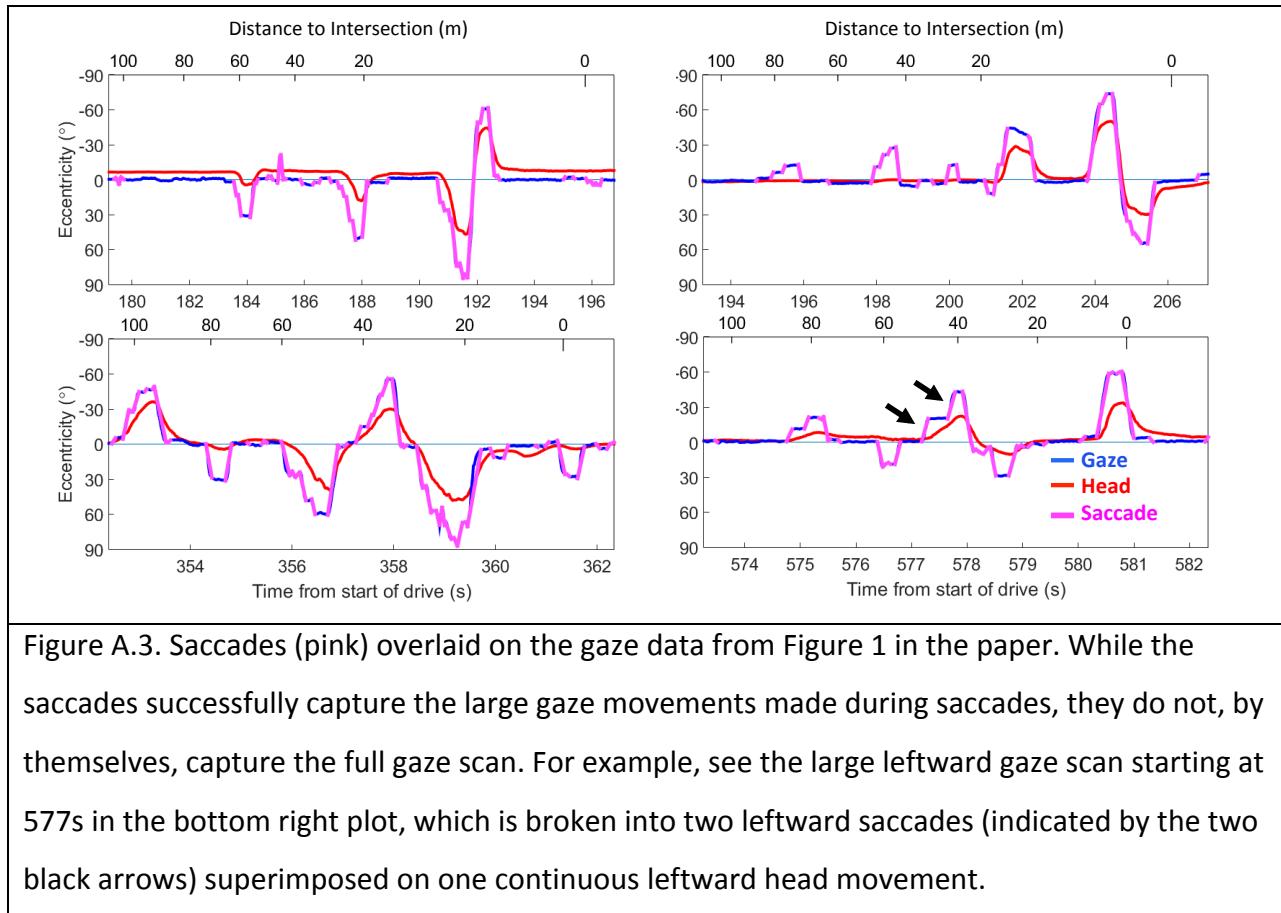


Figure A.3. Saccades (pink) overlaid on the gaze data from Figure 1 in the paper. While the saccades successfully capture the large gaze movements made during saccades, they do not, by themselves, capture the full gaze scan. For example, see the large leftward gaze scan starting at 577s in the bottom right plot, which is broken into two leftward saccades (indicated by the two black arrows) superimposed on one continuous leftward head movement.

776

777

778 A.4. Agreement between manual coders

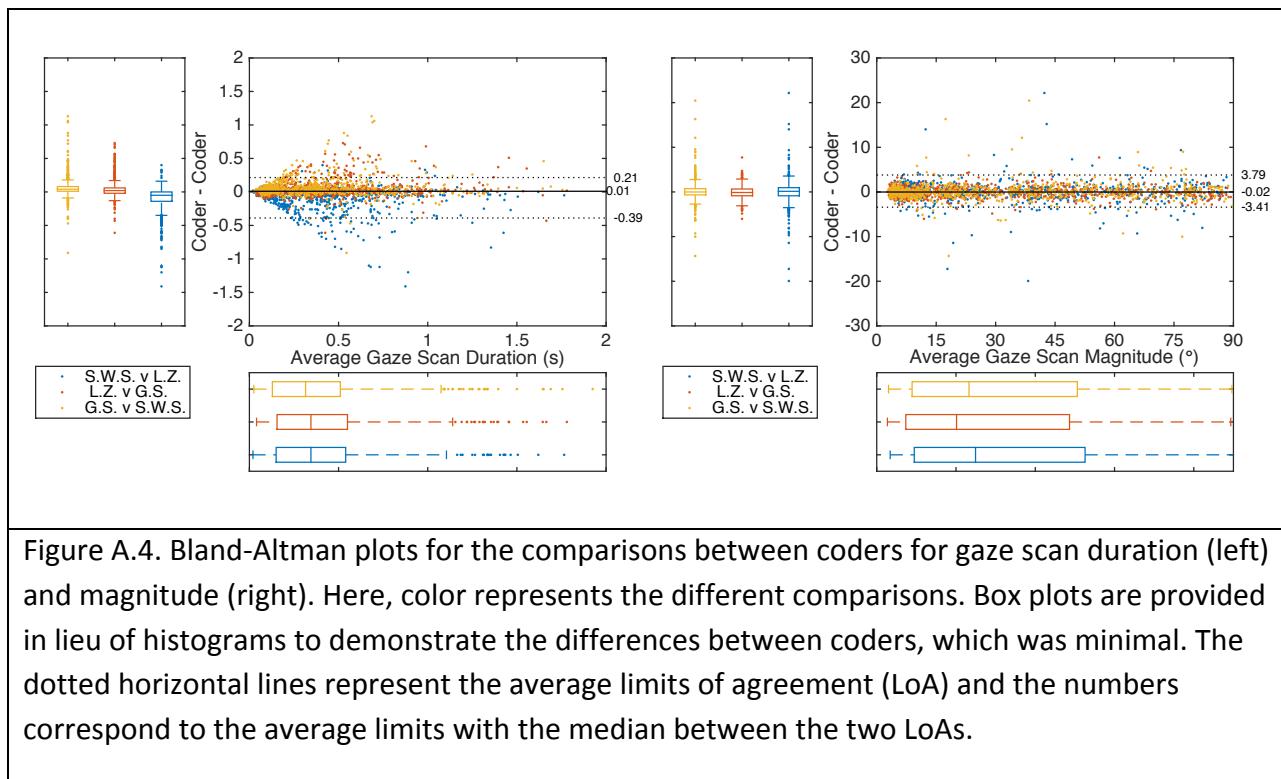
779 Three authors (G.S., S.W S., and L.Z.) independently manually marked data from the
 780 'coders set'. Sample-by-sample Cohen's Kappa between the algorithm and the coders is
 781 displayed in Table A.1. Bland-Altman plots for the comparisons of gaze scan duration and
 782 magnitude are displayed in Figure A.4 and gaze scan start time, end time, start eccentricity, and
 783 end eccentricity are displayed in Figure A.5. As in the main manuscript, the limits of agreement
 784 (LoA) were calculated based on adding the median difference to the 2.5th and 97.5th
 785 percentile, which resulted in an average LoA of 0.30 s (95% CI: 0.24 s to 0.36 s) and 3.6° (95% CI:
 786 3.1° to 4.5°) for duration and magnitude, respectively. In addition, here are the average LoAs
 787 for start time (0.18 s, 95% CI: 0.14 s to 0.22 s), end time (0.21 s, 95% CI: 0.18 s to 0.27 s), start

788 eccentricity (1.72° , 95% CI: 1.43° to 2.1°), and end eccentricity (3.0° , 95% CI: 2.5° to 3.6°). Note,
 789 that there is significant overlap in the markings between coders for each of these measures.

	G.	L.	S.
G.	1		
L.	0.64	1	
S.	0.51	0.68	1
Algorithm	0.71	0.68	0.69

Table A.1. Sample-by-sample Cohen's kappa calculated between the different coders (G.,L.,S.) and between the coders and the algorithm.

790



791

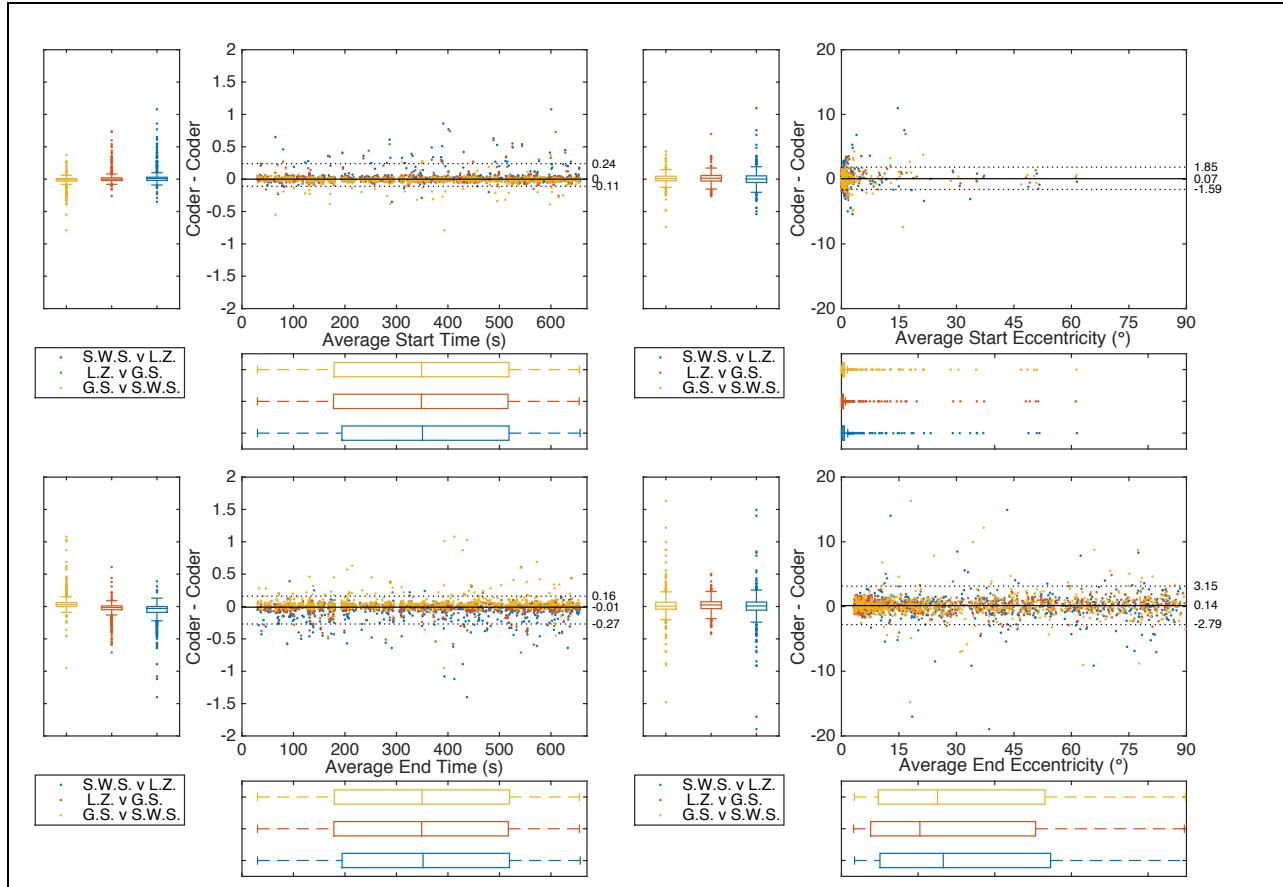


Figure A.5. Bland-Altman plots for the comparisons between coders for gaze scan start time (top left), end time (bottom left), start eccentricity (top right), and end eccentricity (bottom right). Here, color represents the different comparisons. Box plots are provided in lieu of histograms to demonstrate the differences between coders, which was minimal. The dotted horizontal lines represent the average limits of agreement (LoA) and the numbers correspond to the average limits with the median between the two LoAs.