

Manuscript Title: Gaze-in-wild: A dataset for studying eye and head coordination in everyday activities
Author list: Rakshit Kothari, Zhizhuo Yang, Christopher Kanan, Reynold Bailey, Jeff Pelz, Gabriel Diaz

Person	Age	Indoor navigation			Ball catching			Visual search			Tea making		
		Status	Δ_{θ}^{ETG}	Δ_{θ}^{IMU}	Status	Δ_{θ}^{ETG}	Δ_{θ}^{IMU}	Status	Δ_{θ}^{ETG}	Δ_{θ}^{IMU}	Status	Δ_{θ}^{ETG}	Δ_{θ}^{IMU}
1	27	✓✓	0.75	8.71	✓✓	0.67	7.90	×			×		
2	18	✓✓ ⁺	0.79	0.76	✓✓ ⁺	0.89	6.63 ^γ	×			✓	0.76	
3	19	✓	1.27	0.05	✓ ⁺	0.75	2.41	×			×		
6	18	✓	0.43	45.01 ^γ	✓✓	0.37	10.11 ^γ	×			✓ ⁺ <i>O</i>	0.69	0.84
8	22	✓	0.99	4.67	✓	1.29	1.91	✓	0.98	0.56	×		
9	22	✓	0.62	3.66 ^γ	✓	0.73	2.95	<i>O</i>	0.59	0.27	×		
10	26	✓	0.55	13.00	<i>O</i>	1.29	7.36	<i>O</i>	1.04	0.01	<i>O</i>	0.49	0.73
11	25	<i>O</i>	0.77	31.96 ^γ	<i>O</i>	0.56	4.03	×			<i>O</i>	0.69	4.63
12	26	✓	0.52	2.83	✓	0.65	2.04	✓	0.75	1.17	✓	0.57	12.17
13	27	<i>O</i>	1.14	3.70	<i>O</i>	1.08	0.44	<i>O</i>	0.83	7.52	<i>O</i>	0.80	14.14
14	23	<i>O</i>	0.88	0.34	<i>O</i>	1.20	0.02	<i>O</i>	0.82	1.63	<i>O</i>	1.08	1.19
15	22	<i>O</i>	1.09	8.33	<i>O</i>	0.90	1.33	✓	0.73	15.56 ^γ	<i>O</i>	0.67	7.43
16	22	✓	0.82	0.41	✓	0.55	0.53	<i>O</i>	0.94	5.31	<i>O</i>	0.92	0.88
17	23	✓	1.01	4.28	✓	0.89	0.22	✓	0.65	0.22	<i>O</i>	0.79	22.76 ^γ
18	34	✓	1.01	14.46	<i>O</i>	0.99	1.63	✓	0.71	12.09	✓	0.81	3.07
19	26	<i>O</i>	0.88	2.38	✓	0.73	6.52	✓	0.96	19.34 ^γ	×		
20	55	✓	0.93	7.01	<i>O</i>	1.23	3.58	✓	0.95	20.47 ^γ	<i>O</i>	1.24	2.87
22	54	✓	0.50	6.84	✓	0.38	6.89	✓	0.55	11.61	×		
23	60	<i>O</i>	1.23	7.45	<i>O</i>	0.47	3.50	<i>O</i>	0.44	2.04	<i>O</i>	0.44	2.54

Table 1. Dataset status and error measures. $\Delta_{\theta}^{ETG} \rightarrow$ Calibration error in degrees $^{\circ}$. $\Delta_{\theta}^{IMU} \rightarrow$ IMU absolute angular drift measured by the difference in the head orientation at the start and end of a task in $^{\circ}$. Note that certain recordings do not have this measure because we did not explicitly constrain the participants to arrive at their initial pose. For such situations, we estimate head pose when subjects were approximately fixated towards an identifiable direction using visual imagery in the middle of a task. Status symbols: ✓✓ $\rightarrow$ recording has labels from multiple labellers. ✓ $\rightarrow$ recording is labeled by a single person. *O* $\rightarrow$ Not labeled. + $\rightarrow$ Depth values not present. $\gamma \rightarrow$ Manual head pose correction. $\times \rightarrow$ Data discarded.

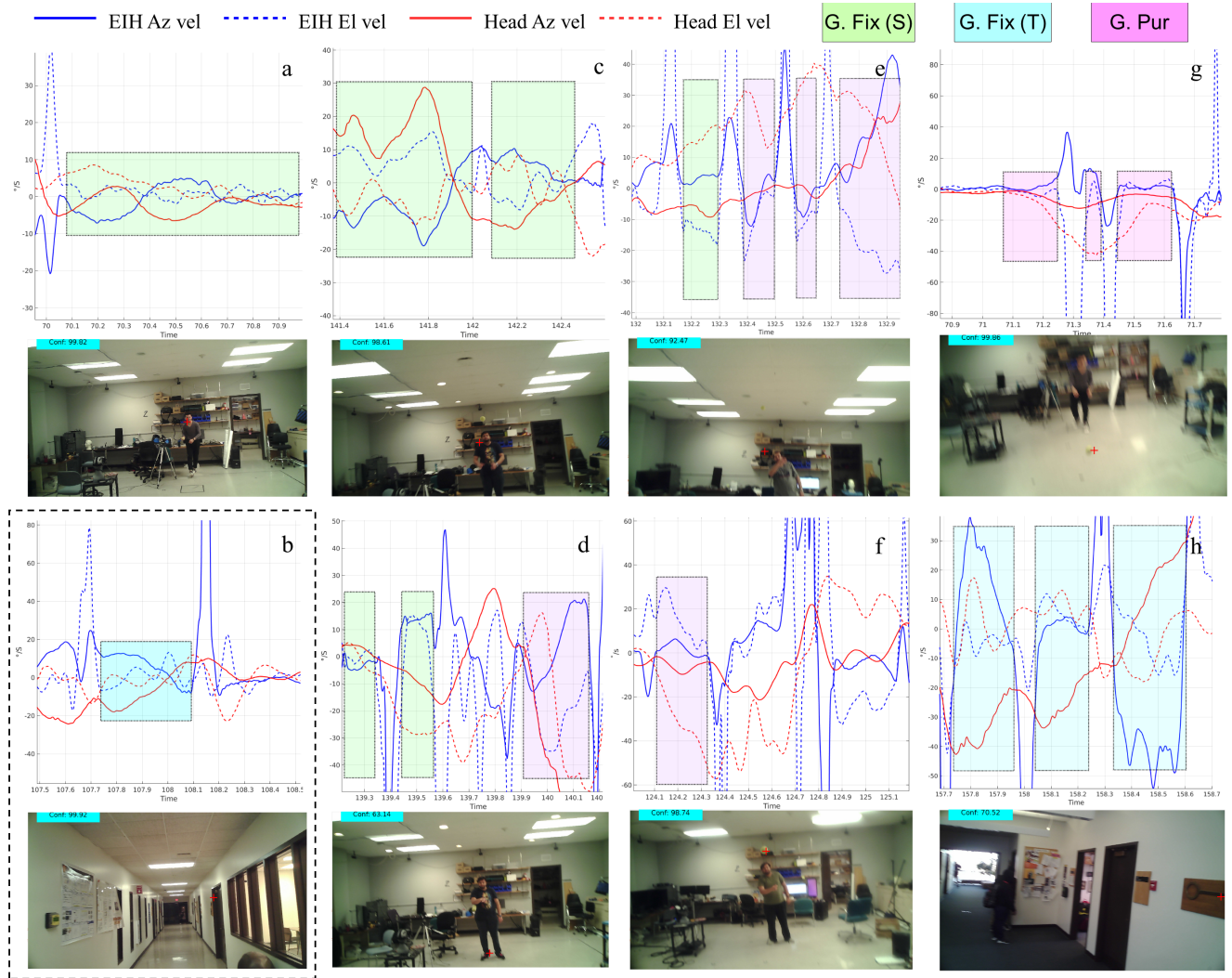


Figure 1. Examples of gaze behavior in 1s clips. Eye and head velocity signals and the related world camera view which best describes the signals is provided as a pair in each row. The bold blue line signifies EiH azimuthal velocity while the dashed line signifies EiH elevation velocity. Likewise, the red variant stands for head velocity traces. Rectangular boxes within each sub-figure are used to draw the readers attention to marked events. Sub-figure **a** and **b** signify gaze fixation events. Note that due to a slight misalignment in the IMU and eye tracker, the velocity components in **a** do not precisely cancel out. Sub-figure **c** and **d** showcase gaze fixation events while the head catches up to the fixation location. Notice that in **d**, the EiH and head velocity traces do not cancel out despite being marked as a fixation. Signal information alone convey that the traces are a head dominated pursuit movement, however, the world imagery shows that the head is catching up to the fixation location. Sub-figures **e**, **f** and **g** highlight examples of head dominated pursuit behavior punctuated by saccades. Note the fixation event marked in **e**. It appears that the eye and head signals compensate in the elevation component but head movements dominate in the azimuthal direction. In hindsight, it appears to be part of a sequence of pursuit events interjected by predictive saccades but has been incorrectly marked as a fixation. Sub-figure **h** is an interesting example of fixation under translation. Visual imagery conveys that the underlying behavior is fixational but one can see that the velocity traces resembles that of a pursuit movement. Examples listed in **d** and **h** contribute heavily to fixation/pursuit confusion.

	Metric	biRNN	biRNN (only eyes)	biRNN (only absolute)	fRNN	RF
Sample level metrics	Overall κ	0.61	0.56	0.58	0.54	0.63
	G. Fix κ	0.61	0.55	0.57	0.54	0.63
	G. Pur κ	0.37	0.24	0.33	0.29	0.28
	Sac κ	0.69	0.71	0.71	0.68	0.74
Event level metrics	Overall (κ^* , κ)	0.37*, 0.47	0.38*, 0.44	0.40*, 0.47	0.27*, 0.35	0.37*, 0.32
	G. Fix (κ^* , κ)	0.34*, 0.61	0.32*, 0.56	0.36*, 0.59	0.21*, 0.49	0.31*, 0.46
	G. Pur (κ^* , κ)	0.14*, 0.27	0.08*, 0.06	0.17*, 0.28	0.03*, 0.12	0.03*, 0.01
	Sac (κ^* , κ)	0.44*, 0.67	0.52*, 0.71	0.51*, 0.70	0.44*, 0.63	0.54*, 0.63
	G. Fix F_1	0.80	0.79	0.80	0.74	0.74
	G. Pur F_1	0.35	0.25	0.34	0.22	0.26
	Sac F_1	0.83	0.85	0.85	0.81	0.82
	G. Fix l_2 (μ , σ)	14.93, 3.64	15.11, 3.46	15.13, 3.65	15.09, 3.32	12.97, 2.10
	G. Pur l_2 (μ , σ)	14.25, 3.84	13.34, 3.75	12.99, 4.90	15.70, 3.44	15.15, 3.44
	Sac l_2 (μ , σ)	14.73, 3.62	14.65, 3.52	14.87, 3.73	15.01, 3.42	12.80, 2.16
	G. Fix O_r	0.92	0.91	0.91	0.90	0.93
	G. Pur O_r	0.90	0.91	0.91	0.89	0.88
	Sac O_r	0.71	0.71	0.71	0.69	0.75
	G. Fix de	0.21	0.20	0.18	0.14	0.09
	G. Pur de	0.20	0.06	0.14	0.13	0.02
	Sac de	0.05	0.05	0.06	0.06	0.11

Table 2. Ablation results reported using all metrics. Note that event level κ^* scores and ratio of detached events (de) are computed using the ELC matching technique as described in Section 4. Mean l_2 distance and it's deviation, $l_2 - \sigma$ between matched events are reported in *ms*.