pEYES: An Open-Source Python Toolkit for Eye Tracking Data Analysis and Detector Evaluation

Jonathan Nir[1], Leon Y. Deouell[1,2]

[1] Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel
[2] Department of Psychology, The Hebrew University of Jerusalem, Jerusalem 9190501, Israel

Introduction

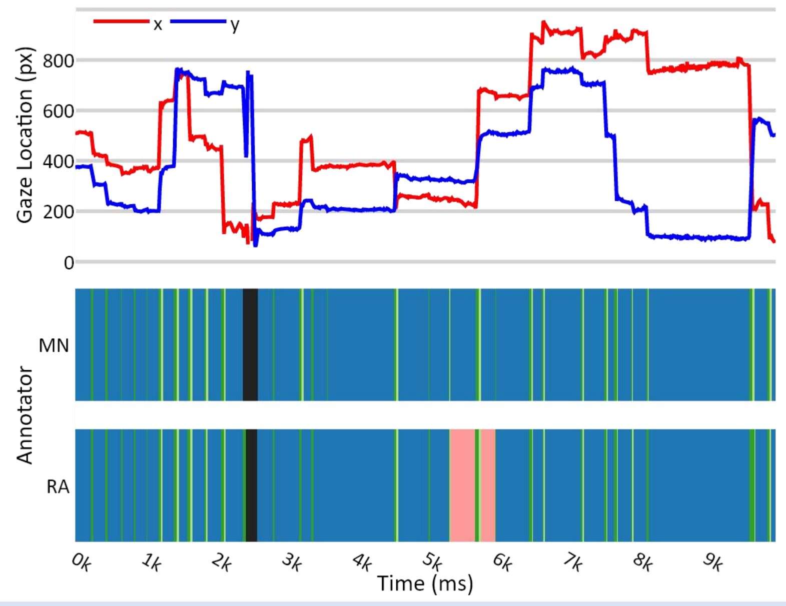
- Eye-Tracking (ET) provides insights into diverse cognitive processes.
- Manually parsing raw ET signal into discrete eye movements (EMs) is difficult and time-consuming.
- Automated algorithms, *detectors*, vary significantly in method, sensitivity, and accuracy.

This Study

- We introduce *pEYES*, a Python package for EM detection and detector evaluation.
- We use *pEYES* to benchmark seven threshold-based detectors on two human-annotated datasets.

The *pEYES* Package

- Standardizes implementations for seven detectors.
- Simplifies EM analyses and visualizations.
- Enables qualitative and quantitative evaluation of detection performance, compared to a ground truth (GT).



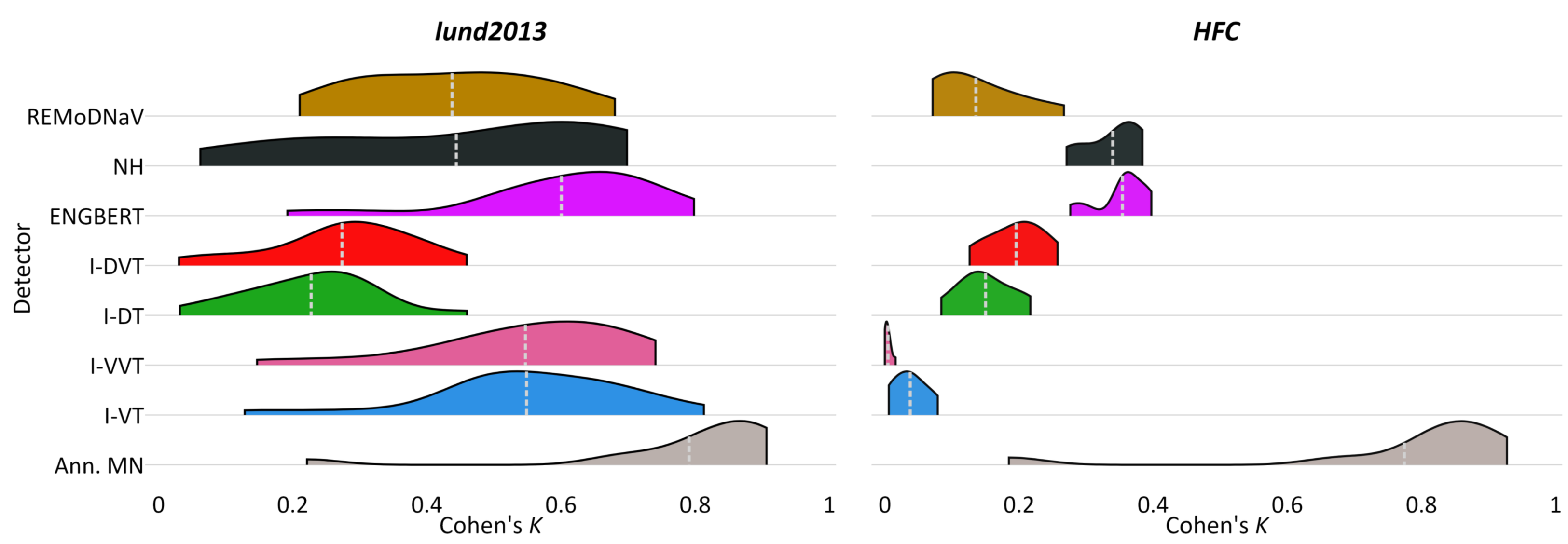
Benchmarking Datasets

	lund2013 [1]	*HFC* [2]
Tracker	iView Hi-Speed 1250	Tobii TX-300
Sampling Rate	500Hz‡	300Hz
Size	87,790 Samples (20 Recordings)	45,018 Samples (10 Recordings)
Stimuli	Color Images	Color Images
Event Types‡‡	U/D, FIX, SAC, PSO, SP, BLI	U/D, FIX
Annotators	**R.A.**, M.N.	**R.A.**, M.N. +10 Others

‡ Four recordings (7,985 samples; 9.1%) were recorded at 200Hz.
‡‡ U/D = Undefined; FIX = Fixation; SAC = Saccade; PSO = Post-Saccadic Oscillation; SP = Smooth Pursuit; BLI = Blink

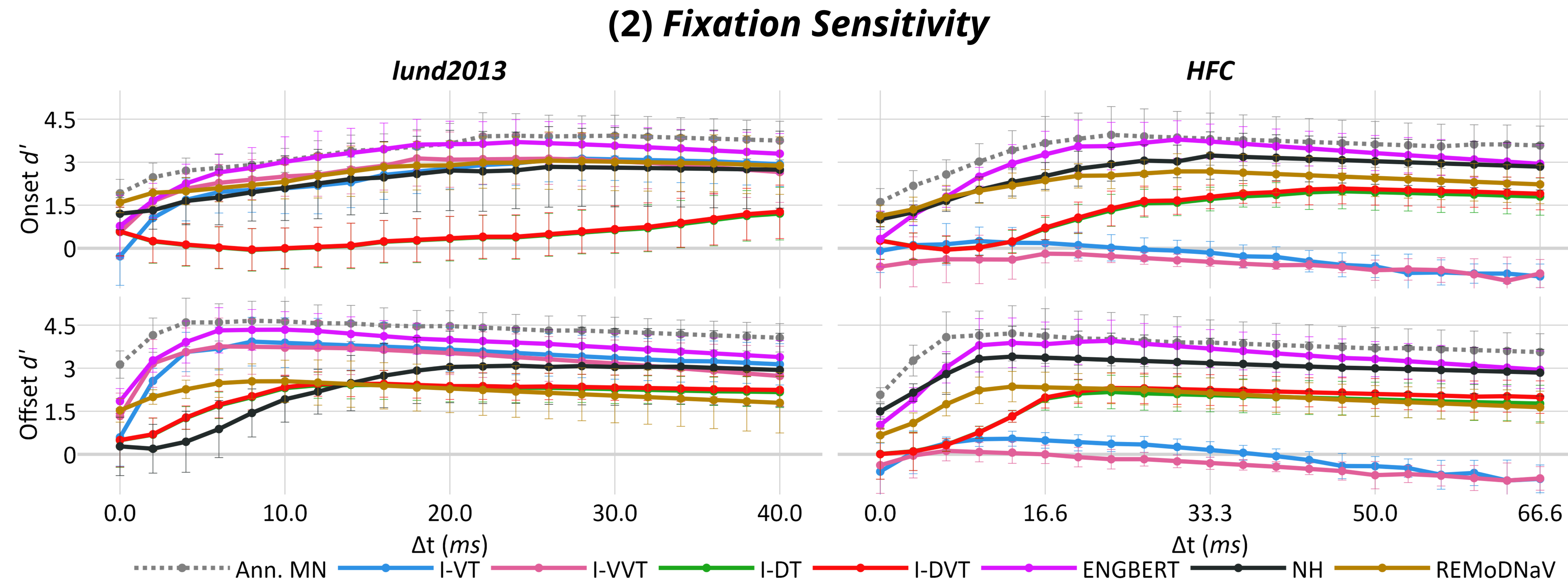
Results

(1) *Agreement with Ground Truth*



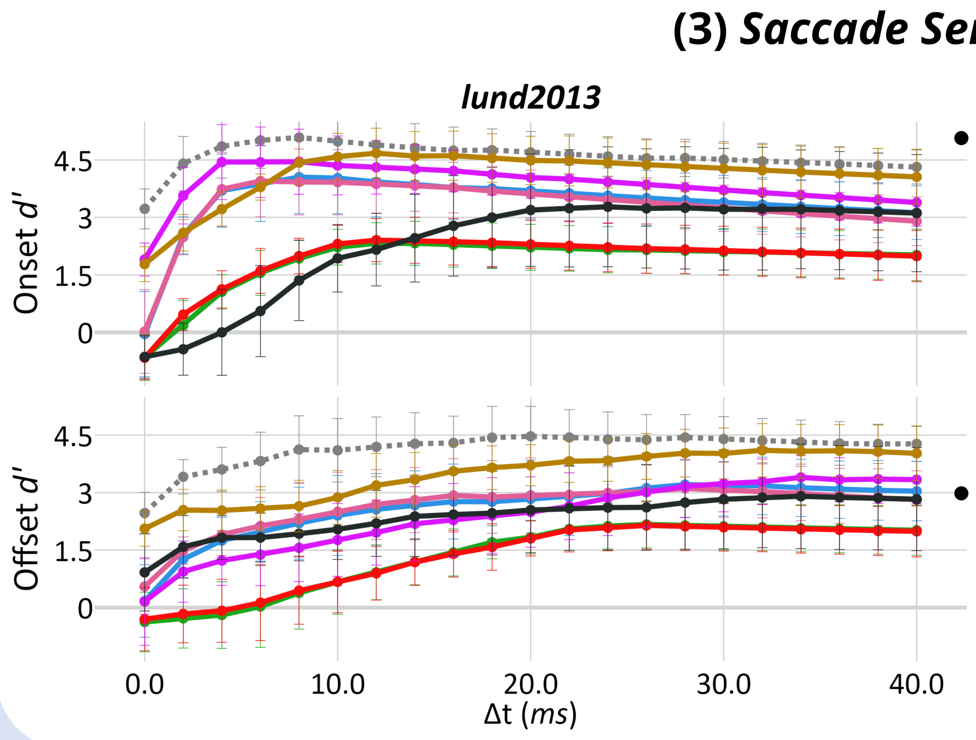
- High sample-by-sample agreement between human annotators, for both datasets.
- Agreement significantly differs across detectors ($p < 0.001$).
- Engbert's algorithm ([3]) matches or outperforms other detectors.

(2) *Fixation Sensitivity*



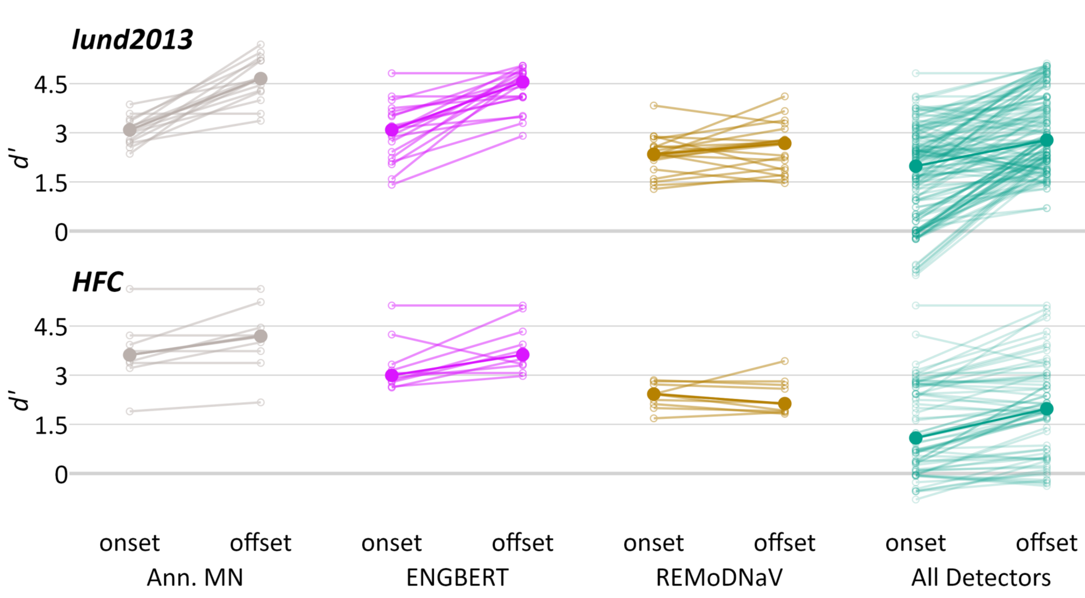
- Engbert's algorithm achieves human-level sensitivity across time windows, for both datasets.
- Global-threshold algorithms (e.g., I-VT, I-DT) perform well on one dataset but generalize poorly to the other.

(3) *Saccade Sensitivity*



- REMoDNaV ([4]) and Engbert's algorithms reach human-level **onset** sensitivity across time windows.
- REMoDNaV outperforms all other algorithms in **offset** detection, across time windows.

(4) Fixation *Onset-Offset Comparison*



- Across datasets and detectors, fixation onsets are more difficult to detect than offsets.
- Saccades present the complement pattern.

Conclusions

1. Detection performance varies significantly between different detection algorithms.
2. The Engbert and REMoDNaV detectors are the optimal algorithm for detecting fixations and saccades during free viewing of static image stimuli, respectively.
3. Fixation onsets are significantly harder than to detect than offsets (vice versa for saccades).
4. *pEYES* provides a framework for assessing detectors in different settings, given reference annotations.

References

[1] Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., & Nyström, M. (2017). One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior Research Methods*.
[2] Hooge, I. T., Niehorster, D. C., Nyström, M., Andersson, R., & Hessels, R. S. (2018). Is human classification by experienced untrained observers a gold standard in fixation detection?. *Behavior Research Methods*
[3] Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*.
[4] Dar, A. H., Wagner, A. S., & Hanke, M. (2021). REMoDNaV: robust eye-movement classification for dynamic stimulation. *Behavior Research Methods*.
[5] Startsev, M., & Zemblys, R. (2023). Evaluating eye movement event detection: A review of the state of the art. *Behavior Research Methods*.

ISRAEL SCIENCE FOUNDATION
THE ISRAELI SMART TRANSPORTATION RESEARCH CENTER