pymovements: Building Blocks for Reproducible Eye-Tracking Pipelines

Daniel G. Krakowczyk[*1,2], **Carlson M. Büth**[1], **David R. Reich**[1,2], **Deborah N. Jakobi**[1], **Paul Prasse**[2], **Andreas Säuberli**[3], **Anastassia Shaitarova**[1], **Lena A. Jäger**[1]

[1]Department of Computational Linguistics, University of Zurich, Switzerland
[2]Department of Computer Science, University of Potsdam, Germany
[3]Center for Information and Language Processing, LMU Munich, Germany

*contact: danielgeorg.krakowczyk@uzh.ch

pymovements is an open-source Python package for processing and analyzing eye-tracking data, designed to address fragmentation in preprocessing and analysis practices. Preprocessing pipelines are frequently implemented independently across laboratories, resulting in inconsistent filtering, event detection, and data handling procedures. This variability reduces reproducibility, complicates cross-study comparisons, and increases barriers for new researchers. *pymovements* mitigates these issues by providing a unified, transparent, and extensible framework that standardizes core workflows. The project has been in active development since 2022 and is closely linked to the MultiplEYE COST Action as a foundation for its data preprocessing pipeline.

The package supports parsing of widely used eye-tracking formats, including EyeLink and BeGaze, and provides access to more than 35 publicly available eye-tracking datasets. It includes functionality for filtering and transformation, data quality assessment and data cleaning, as well as detection of oculomotor events such as fixations, saccades, and blinks. Together, these components establish a consistent basis for downstream analysis.

On top of these core primitives, *pymovements* provides tools for computing derived measures used in psycholinguistic and reading research, including stimulus-aligned analyses for textual data and standardized visualization tools such as heatmaps, scanpaths, and saccadic main-sequence plots, facilitating both exploratory data analysis and methodological evaluation.

The package is developed under continuous integration with automated testing, mandatory code review, and comprehensive documentation. It includes initial support for BIDS-compliant Gorgolewski *et al.* (2016) formats and adheres to FAIR principles Wilkinson *et al.* (2016), promoting findability, accessibility, interoperability, and reusability of eye-tracking data and analysis workflows. Released under the MIT License, *pymovements* is developed collaboratively by a distributed community of more than 45 contributors from multiple research laboratories. We actively encourage contributions that may take the form of documentation improvements, dataset integrations, bug reports, and feature enhancements.

Overall, *pymovements* provides a modular and extensible infrastructure for reproducible eye-movement research. By integrating essential components of eye-tracking analysis into a coherent and standardized framework, it reduces methodological heterogeneity, improves transparency, and supports robust, reproducible research practices across the field.

References

Gorgolewski K. J., Auer T., Calhoun V. D., Craddock R. C., Das S., Duff E. P., Flandin G., Ghosh S. S., Glatard T., Halchenko Y. O., Handwerker D. A., Hanke M., Keator D., Li X., Michael Z., Maumet C., Nichols B. N., Nichols T. E., Pellman J., Poline J.-B., Rokem A., Schaefer G., Sochat V., Triplett W., Turner J. A., Varoquaux G., Poldrack R. A. (2016, June). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data 3*(1). doi:10.1038/sdata.2016.44.

Wilkinson M. D., Dumontier M., Aalbersberg I. J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.-W., da Silva Santos L. B., Bourne P. E., Bouwman J., Brookes A. J., Clark T., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C. T., Finkers R., Gonzalez-Beltran A., Gray A. J., Groth P., Goble C., Grethe J. S., Heringa J., 't Hoen P. A., Hooft R., Kuhn T., Kok R., Kok J., Lusher S. J., Martone M. E., Mons A., Packer A. L., Persson B., Rocca-Serra P., Roos M., van Schaik R., Sansone S.-A., Schultes E., Sengstag T., Slater T., Strawn G., Swertz M. A., Thompson M., van der Lei J., van Mulligen E., Velterop J., Waagmeester A., Wittenburg P., Wolstencroft K., Zhao J., Mons B. (2016, March). The fair guiding principles for scientific data management and stewardship. *Scientific Data 3*(1). doi:10.1038/sdata.2016.18.