One man’s trash is another man’s treasure: The need of open data and tools for research reproducibility in screen-based eye-tracking marketing experiments

EMAC 2024 – Bucharest, Romania

Vasileios Mygdalis and Nathalie Dens
Department of Marketing, University of Antwerp

Abstract

Despite the existence of a wide range of eye-tracking experiments in marketing studies, deriving cross-study consensus remains a challenging task. In this article, we identify and discuss some common causes of inter-study variance and propose a set of tools for minimizing cross-study inconsistencies. More precisely, we discuss expected cross-study variances related to how eye-tracking equipment is used, while we also identify other sources of cross-study variance related to picking different eye-tracking variables, differences in the procedure of defining AOI, differences in the experimental protocol and data processing. Based on this review and inspired by the best practices that have been followed in the computer vision discipline leading to substantial advancements in the past few years, we propose a set of practical and methodological tools that could be taken into consideration for eye-tracking data quality reassurance, maximizing individual study reproducibility and assist in reaching cross-study consensus.

Inter-study differences in eye-tracking experiments in marketing

Different experimental protocols

- Eye-tracker precision fluctuates among participants.
- Scene geometry and lighting conditions matter.
- Distance from the screen, screen size and visual stimuli sizes significantly affect the sensor spatial resolution.
- *S*ensor calibration per subject has a severe impact on sensor precision and recall. Appropriate calibration per subject is not always trivial, especially when dealing with non-adults.

Inconsistencies in AOI definition

- AOI definition typically involves defining rectangular objects, while the visual stimuli may not be rectangular.
- Some researchers opt for increased AOI sizes, to compensate for sensor inaccuracies in fixation detection. This is wrong. It has been proven by recent research that increasing the AOI size maybe increases fixation recall by with a cost in precision.

Plethora of naming conventions and ill variable definition

- Total fixation time refers to the time the participant spent viewing the stimuli, adding fixations detected by some algorithm, by performing analysis of the raw signal (gaze points). Some studies add fixations of a minimum duration of 58ms, 80ms, others 150ms, others 300ms, or others do not report which variation was used.
- Same variables can have different naming conventions. Total fixation time might be referred as fixation duration, consumer attention, fixation, eye-tracking.
- Many different variable variations are used without providing any additional insights, like average fixation duration. The added value of selecting such variations is questionable.

Data analysis prototyping

Data acquisition quality assurance

- How to we know if the eye-tracking data we are collecting are enough and correct? We propose the introduction of metrics :
- **Inter-observer-agreement** -> how similar are the attention maps of different participants?
- **Saturation** -> the exact number of study participants that lead to stabilization of geometric means in the core study variables
- **cross-database similarity ->** Do my data look similar with other publicly available databases?

Data sharing

- Sharing a dataset of eye-tracking research is very easy, the incentives to this end are huge, yet it does not occur. Why?
- The value of raw data is under-appreciated.
- Proprietary software is used for the analysis.
- Open data formats although exist, they are not always used.

Open data processing and analysis tools

- Webcam-based eye-trackers are in many cases accurate enough. Do we really need high refresh rates?
- Raw sensor data can be analyzed with open-source analysis tools (such as PyGaze). The algorithms for calculating fixation time, scan path, time to first fixation are implemented there.
- The statistical analysis can be conducted in Python. There is no need for more specialized software.

Reproducibility incentives and considerations

- Dataset papers in other domains are typically highly cited.
- The raw gaze data acquired during a marketing experiment could be re-used in other studies and other domains.
- In case raw gaze data sharing is not an option, the visual stimuli and the extracted core variables (fixation time, heatmaps, scan paths) should still be shared, even if they are irrelevant to the study.
- Highly specialized, processed variables/aggregations or ill-defined variables cannot be re-used in other studies.

Conclusion

- Many of the above-mentioned issues can be alleviated by simply promoting and enforcing to some extent, research reproducibility considerations.
- We promote open-source analysis tools, streamlining the analysis starting from proper stimuli definition, raw signal breakdown to statistical analysis and even deep learning.

Universiteit Antwerpen

CONVISE project has received funding from the European Union’s HE research and innovation programme under the g.a.n. 101103256

www.convise-project.eu

CONVISE
CONSUMER VISUAL ATTENTION SENSING