

# Tracking Non-Visual Eye Movements Non-Invasively: Comparing Manual and Automatic Annotation Styles

Jeremias STÜBER, Lina JUNCTORIUS, Annette HOHENBERGER

Institute of Cognitive Science, University of Osnabrück,  
49069 Osnabrück Germany

{jstueber, ljunctorius, annette.hohenberger}@uni-osnabrueck.de

**Abstract.** Non-visual eye-movements (NVEMs) during episodic and semantic memory retrieval may reveal mental foraging, similar to bodily movements during physical foraging. However, measuring them is challenging as common eye-tracking methods might interfere with the natural eye gaze habits of participants. Our aim was to compare two annotation approaches to track NVEMs from raw video footage: First, manual annotation using a coding grid dividing the visual field into nine sections, and second, automated annotation using the neural-network driven face recognition software OpenFace representing eye gazes as Cartesian vectors. We found that both approaches showed moderate to excellent reliability, quantitatively. *k*-mean clusterings of the automatic annotations revealed high resemblance to the manual coding approach, qualitatively. We conclude that mapping between manual and automatic coding of NVEMs is feasible, and that the selection of the appropriate method should depend on the research question and available resources.

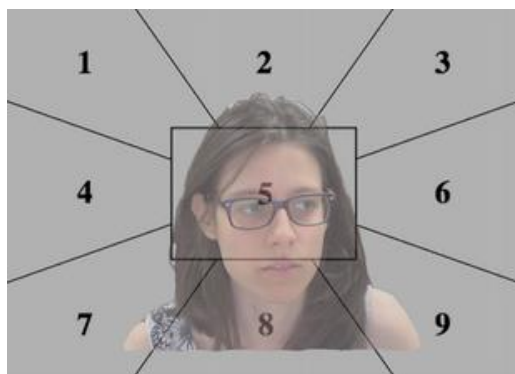
**Keywords:** Eye movements, non-visual, methodology, annotation

Non-visual eye-movements (NVEMs) are eye-movements that do not serve the provision of visual information but are related to memory retrieval processes (Ehrlichman and Micic, 2012; Micic et al., 2010). A commonly held belief, purported by adherents to neuro-linguistic programming (NLP), is that certain eye gaze directions are indicative of certain cognitive states, e.g., remembering, imagining, or having an internal dialog (Florea et al., 2013). Even though NLP has been subject to criticism from the scientific community, their claims point towards an area of research that has not yet been explored in much detail, nor with methodological rigor. Another conjecture is that the way our internal memory search unrolls is similar to actual (foraging) movements in physical space, e.g., when naming category members similar ones in mental space can be remembered faster whereas dissimilar ones can only be accessed after searching further, according to an exploit/explore process (Hills and Butterfill, 2015; Patten et al., 2020). This literature suggests that physical or mental foraging of a patchy resource follows a Lévy-Flight, or other “heavy-tailed” distribution. Inspired by this research, here we first lay the method(olog)ical foundations for studying NVEMs as evidence for mental

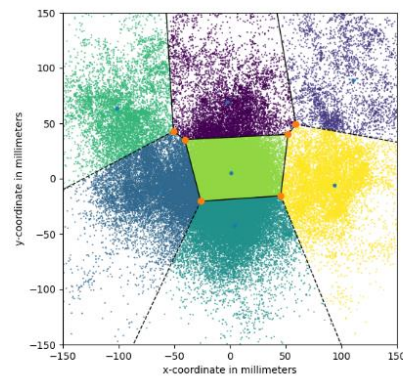
foraging in another area of cognition, namely episodic and semantic cognition. Participants had to remember and report episodes of the past or imagine future episodes related to or providing a semantic description of an object, e.g., a door or key.

A basic difficulty that arises in investigating correlations between cognitive processes and NVEMs is one of annotation. Due to the nature of NVEMs, eye gazes are not bound to a specific surface or location as they are during eye-tracking tasks. Therefore, conventional eye-tracking paradigms are not apt to deal with NVEMs. Furthermore, head-mounted devices, due to their disruptive nature, are likely to interfere with a participant's normal NVEMs habits, distorting the data as a result. Here, we present and discuss two approaches that operate on the basis of face-centered video footage of participants, obtained from an external camera positioned above the computer screen. The first approach involves manual annotation in the open-source application ELAN (Elan, 2020) according to a coding grid which divides the visual field into nine sections (plus one "other"), in the form of "pie slices" around a central region where, on the computer screen, pictures of the to be mentalized objects were presented (see Fig. 1a). Inter-rater reliability between two manual annotators was excellent (Cohen's  $\kappa = 0.927$ ).

The second approach makes use of the open-source neural-network driven face recognition software OpenFace (Baltrusaitis et al., 2018), allowing for the representation of the participant's NVEMs in terms of Cartesian vectors. After preprocessing involving (i) coordinate system rotation (since each participant looked at the center of the screen from a slightly different angle and position), and (ii) removing the depth dimension (z-axis) from the data by means of a planar projection, we calculated the overall agreement of vectors for each grid cell using the intraclass correlation coefficient (ICC), with participants as "raters". The ICC for the (horizontal) x-dimension was 0.83 (95% CI: 0.71-0.94) and for the (vertical) y-dimension 0.74 (95% CI: 0.58-0.90), indicating a 'moderate' to 'good' reliability (Koo and Li, 2016). These ICC estimates give reason to believe that an automated analysis with OpenFace is indeed viable. The final step in our analysis of the OpenFace data consisted in determining whether the coding grid could be reconstructed using a *k*-means clustering algorithm. The clearest partitioning of the data



**Figure 1a)** Manual coding grid with participant gazing in direction "6" (from the perspective of the camera).



**Figure 1b)** *k*-means clustering of OpenFace vector-based gaze points (data collapsed from all participants across all conditions, from the perspective of the participant).

was obtained with 7 clusters (see Figure 1b):

The reconstructed clusters resemble the manual coding grid to a high degree, confirming the good quantitative match in terms of the ICC also qualitatively. The central region (corresponding to “5”) is very prominent: it is the “anchor” area from which eye gazes depart and to which they return as the object is displayed there. The two grids differ most laterally, where the upper (1 and 3) and lower slices (7 and 9) incorporate the middle slices (4 and 6), effectively reducing  $k$  from 9 (manual coding grid, Fig. 1a) to 7 (OpenFace vector-based grid, Fig. 1b).

Given the good quantitative ICC score and the qualitative resemblance of the manual and the reconstructed vector-based grid, we conclude that a mapping from one representation to the other is feasible and hence both approaches work. Regarding the superiority of one approach over the other, deciding this particular question should in part depend on the research question one tries to investigate, as well as the resources one has access to. Where the manual approach requires repetitive, costly manual labor, OpenFace requires some degree of proficiency with programming and data visualization techniques but is cheap, once mastered. As an additional virtue, OpenFace has performed very well in comparative studies with other eye/face-tracking software (Baltrusaitis et al., 2018), and seems to be less biased in terms of the assumptions it makes about human eye-gazes.

Overall, our chief finding is best described as providing two distinct, but potentially converging approaches to the study of NVEMs. Having settled this important method(ological) issue, we now avail over tools through which we can study our original question in what way NVEMs are associated with our cognitive processes, most notably, whether the parallelism between “perceptual foraging” (through physical eye movements in space) and “mental foraging” (memory search) actually holds. If so, this would add a further facet to the study of spatial cognition.

## Acknowledgements

We wish to thank Jana Kernos for designing the original experiment.

## References

- Baltrusaitis, T., Zadeh, A., Chong Lim, Y., Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit. *13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 59-66 - May 2018. <https://doi.org/10.1109/fg.2018.00019>
- Ehrlichman, H., Micic, D. (2012). Why do people move their eyes when they think? *Current Directions in Psychological Science*, 21(2), 96-100. <https://doi.org/10.1177/0963721412436810>
- Elan (version 6.0). Computer software, Nijmegen: Max Planck Institute for Psycholinguistics, *The Language Archive*, 2020. <https://archive.mpi.nl/tla/elan>
- Florea, L., Florea, C., Vranceanu, R., Vertan, C. (2013). Can your eyes tell me how you think? A gaze directed estimation of the mental activity. *Proceedings British Machine Vision Conference 2013*. Pages 60.1-60.11 <http://dx.doi.org/10.5244/C.27.60>

- Hills, T.T., Butterfill, S. (2015). From foraging to auto-noetic consciousness: The primal self as a consequence of embodied prospective foraging. *Current Zoology*, 61(2), 368-381. <https://doi.org/10.1093/czoolo/61.2.368>
- Koo, T.K., Li, M.Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Micic, D., Ehrlichman, H., Chen, R. (2010). Why do we move our eyes while trying to remember? The relationship between non-visual gaze patterns and memory. *Brain and Cognition*, 74, 210-224. <https://doi.org/10.1016/j.bandc.2010.07.014>
- Patten, K.J., Greer, K., Likens, A.D., Amazeen, E.L., Amazeen, P.G. (2020). The trajectory of thought: Heavy-tailed distributions in memory foraging promote efficiency. *Memory & Cognition*, 48, 772-787. <https://doi.org/10.3758/s13421-020-01015-7>

Received July 12, 2021, accepted August 25, 2021