Using Pupillometry as a Surrogate for Brain Reward Feedback with Commodity Hardware and Software

Martin Diges (mdiges@wisc.edu)
Department of Computer Science, 1210 W. Dayton Street
Madison, WI 53706 USA

Abstract

Prior work suggests that measurements of pupil diameter, pupillometry, can reveal some information about an individual's inner mental state, a prominent example of which being information about recent rewards. Can measures of pupil diameter captured using commodity hardware and software be used as proxies for reward feedback as it is perceived in the brain?

Keywords: pupillometry; reward signaling; commodity hardware

Background and Motivation

A Brain Computer Interface (BCI), is a type of human-machine interface that reads signals from the brain, typically electric signals, and translates them into desired actions. These devices are developed in a clinical setting most often for persons with significantly reduced motor control and can allow those persons to control a wheelchair, use a computer, or plainly just communicate.

A BCI must be learn to interpret a user's brain signals in order to help them perform tasks appropriately. One way to teach a BCI is through Supervised Learning (SL), where the user is exposed to stimuli (which can include proprioceptive stimulus) that elicit brain signals associated with their intent to perform a task. The BCI makes a prediction and compares its prediction to what the user's "true intent" is. Suminski et al. (2010) demonstrate an instance of this Supervised Learning in the context of training a BCI to move a robot arm on behalf of a nonhuman primate. SL requires a "ground truth" against which the BCI may compare their prediction, hence a dataset is needed of "true" robot arm movements. This approach poses an important issue, which is that such "ground truth" data may not be readily available. This hampers the autonomy that BCIs are meant to provide, as the persons using them may require controlled training for each new task they wish their BCI to perform.

Reinforcement Learning is another approach to training BCIs. Rather than compare its prediction to a "true intent", the BCI receives feedback on how good its prediction is through some other mean. The advantage with using RL is that the BCI only needs a reward signal as opposed to ground truth data. If a person has the ability to specify the reward correspondent to the BCI's predictions, the BCI can improve at their task over time without the need for a special dataset, a controlled training environment, or a visit to a facility.

The final issue that remains is that individuals with heavily reduced motor function may have trouble providing this feedback to the BCI.

In their review of reinforcement learning BCI approaches, Girdler, Caldbeck, and Bae (2022) highlight existing work, including An et al. (2019), which provide evidence for obtaining such a reward signal directly from a person's brain. Thus, the thought goes, if reward can be obtained from the person's brain, they need only be able to assess the BCI's performance at a mental level in order for the BCI to learn.

Predicting reward from brain signals works for BCIs with access to the inside of the skull, but what if one would like to obtain a reward signal without having to place electrodes on/inside the skull? Pupillometry, the analysis of pupil diameter dynamics, is one such approach. Work by Kloosterman et al. (2015) suggests that changes in pupil size at constant light levels may be reflective of some brain state, and in particular that pupil dynamics can encode some of the surprise and content of perceptual events.

Question

If pupil diameter can carry some information about events being perceived, could pupillometry be used as a surrogate for brain reward feedback?

Method

A microphone stand was modified and set up such that it could hold a camera and so that it could be moved close to a participant's face when desired. To provide a live video feed, a Google Pixel 8 phone was placed on the stand, connected to a computer via a stock USB-C cable (with data transfer capabilities), and it was made to act as a webcam through its USB settings. This provided a significantly superior image quality compared to a regular webcam, which doesn't fare well when close to one's face.

On the same computer, a simple GUI program was created in Python 3.13 using the built-in "tkinter" GUI library. A screenshot of this GUI program is shown in Figure 1. The GUI displays the camera feed to the user. The program uses OpenCV code to detect eyes in the camera feed's frames. Further code locates the likely pupil within the eye area. The predicted pupil diameter is overlaid on the camera feed for the user's convenience. The GUI also plots pupil diameter over time as diameter readings come in.

Task

Finally, the GUI displays a simple 2-alternative forced choice (2AFC) game for the user to play. Two shapes are present side by side. During a single trial, the player may choose between the two shapes by pressing the arrow keys on their keyboard. One of the shapes is the "correct" shape and will

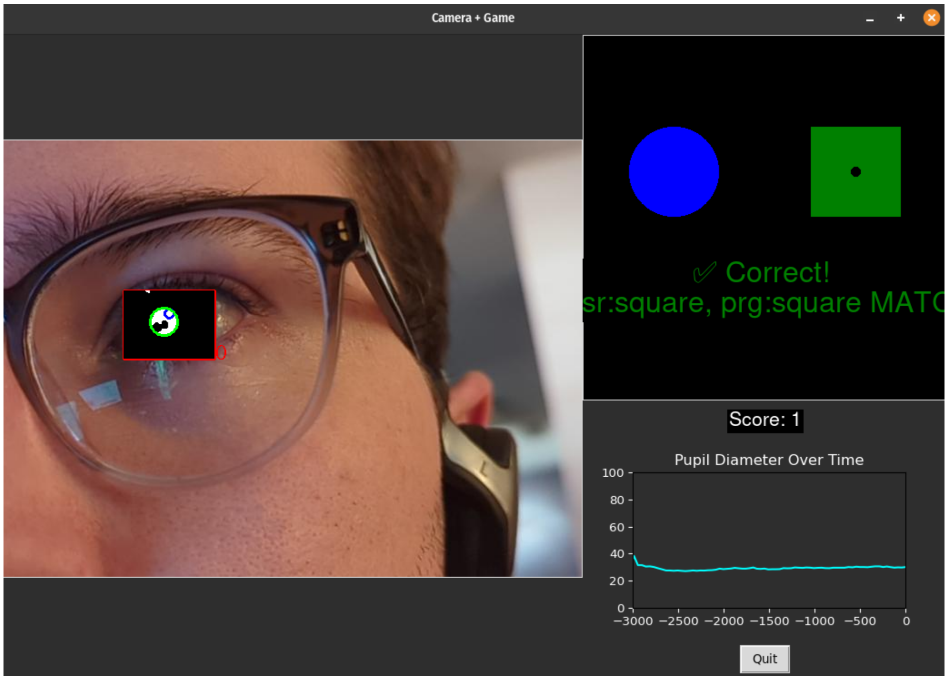

Figure 1: Screenshot from the GUI program during operation

inform the player of their correct choice visually and with a pleasant auditory queue. The other shape provides corresponding visual feedback on the player's incorrect choice and also provides a less pleasant auditory queue. The visual feedback is kept small so as to try to avoid pupil dynamics caused by changes in screen luminosity. As the player chooses correctly, the "correct" shape may change with some probability after a successful trial.

The user's pupil diameter is recorded for the duration of the GUI being open and is written to disk. When analysis begins, the pupil trajectories for each of the trials are aggregated within a range of time around the time of the user being rewarded. This range is (-500, +1000) ms so as to capture some of the steady-state pupil diameter before reward is provided and how the pupil diameter evolves after the reward is provided. Figure 2 shows the preprocessing that these traces undergo. Firstly, the L1 norm of the pupil diameter across all trials is calculated. Trials wherein the pupil diameter goes further from the mean than 4 times this norm are considered outliers and are marked for removal. Outliers can pop up as a result of the imperfect pupil diameter detection system used, as it can momentarily lose track of the pupil. Secondly, the trials are coded according to whether the reward administered was positive or negative. Lastly, as is seen in Figure 3 in the results section, pupil diameter samples are binned into 50ms segments of time so that samples can be aggregated over trials.

Results

Figure 3 shows our main results. Once the pupil diameter samples are binned (d), they have their mean plotted with a 95% non-parametric CI (e). There does not appear to be a statistically significant difference between pupil diameter before the reward time.

From (d), it was observed that there was some per-trial bias in the pupil diameter. This is a consequence of the imperfect camera setup where the participant's head was not stationary and hence small head movements could periodically change the steady state pupil diameter. Hence in (f) each trial has its pre-reward mean pupil diameter computed and subtracted from all trial samples. What we see are traces that have much of their per-trial bias removed. Finally, in (g) we take the mean of these "de-meaned" (not literally, of course) trajectories and again plot it with a 95% CI. The confidence intervals once again indicate no statistically significant difference before the reward time. Though not statistically significant at a confidence of 95%, the plot in (g) suggests that there may be a difference between the post-reward pupil diameter of correct/incorrect trials that could be used to predict reward.

Discussion

The type of reward signal being examined during this experiment is not what would be used to train a BCI. A BCI would require feedback on its predictions, whereas the task presented to the player in this paper does not have a model making predictions from user input and thus does not elicit such feedback. The type of reward feedback examined in this paper could most closely be described as the player's perception of their own performance. The machinery to elicit the prior type of feedback was present in the code but went unused primarily owing to how confusing it made the task and the limitation on statistical power described below. Further

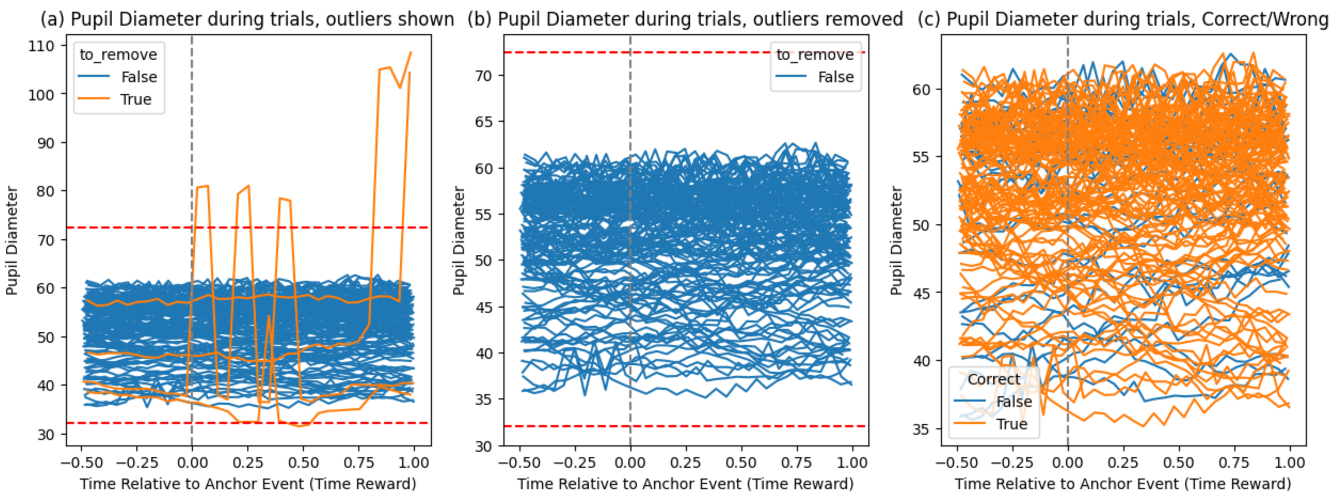

Figure 2: Preprocessing carried out ahead of analysis. Vertical dashed line indicates the time at which reward was administered to the user. **(a)** Pupil diameter samples from the time interval 500ms before to 1000ms after reward is given. Each trace corresponds to the diameter samples from a single trial. Horizontal dashed lines are placed 4 L1 norm distances from the all-trial mean pupil diameter. Traces which exceed this threshold are marked for removal. **(b)** Same information as (a), only with the outlier traces removed. **(c)** Same traces as shown in (b), with the horizontal dashed lines removed and traces now colored according to whether the shape chosen was correct or not.

work would address these two concerns.

The number of trials used after filtering was quite small (n=142), hence my hypothesis is that with a greater number of trials there may be a statistically significant difference in pupil diameter between correct and incorrect trials. This limitation was present due to the inconsistency in steady-state pupil diameter between recording sessions, itself a consequence of the physical setup of the experiment not keeping the subject's head a steady distance from the camera. The consequence of this session-to-session difference in steady-state pupil diameter is that samples cannot readily be aggregated across sessions, hence only samples from a single session could be reliably used. 142 is just about the most trials that could be tolerated while staying still with eyes wide open. It should be possible to address this limitation, but constraints on time leave this for future development.

Conclusion

In this paper we discussed an attempt at differentiating brain reward state through pupil dynamics observed by commodity hardware and software.

While the difference in pupil dynamics was not significant at a 95% level, the overall shape of the intervals does motivate the aggregation of more sessions to see whether a statistically significant difference can be found.

Acknowledgments

This paper was inspired by preliminary findings made while working with faculty and researchers from the University of Wisconsin - Madison's Department of Neuroscience. Thank you to Aaron Suminski, Luis Populin, and Abigail Rajala for their mentorship.

This paper was created for Comp Sci / Psych 841, taught by Joseph Austerweil. Thank you to Joe Austerweil for being an encouraging instructor and for making sure I kept my expectations realistic. Without their feedback I likely would not have finished this work in time.

References

An, J., Yadav, T., Hessburg, J. P., & Francis, J. T. (2019). Reward expectation modulates local field potentials, spiking activity and spike-field coherence in the primary motor cortex. *eNeuro*, *6*(3). Retrieved from `https://www.eneuro.org/content/6/3/ENEURO.0178-19.2019` doi: 10.1523/ENEURO.0178-19.2019

Girdler, C. W. . B. J., B. (2022). Neural decoders using reinforcement learning in brain machine interfaces: A technical review. *Frontiers in systems neuroscience*, *16*(836778). Retrieved from `https://doi.org/10.3389/fnsys.2022.836778`

Kloosterman, N. A., Meindertsma, T., van Loon, A. M., Lamme, V. A. F., Bonneh, Y. S., & Donner, T. H. (2015). Pupil size tracks perceptual content and surprise. *European Journal of Neuroscience*, *41*(8), 1068-1078. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/ejn.12859` doi: https://doi.org/10.1111/ejn.12859

Suminski, A. J., Tkach, D. C., Fagg, A. H., & Hatsopoulos, N. G. (2010). Incorporating feedback from multiple sensory modalities enhances brain–machine interface control. *Journal of Neuroscience*, *30*(50), 16777–16787. Retrieved from

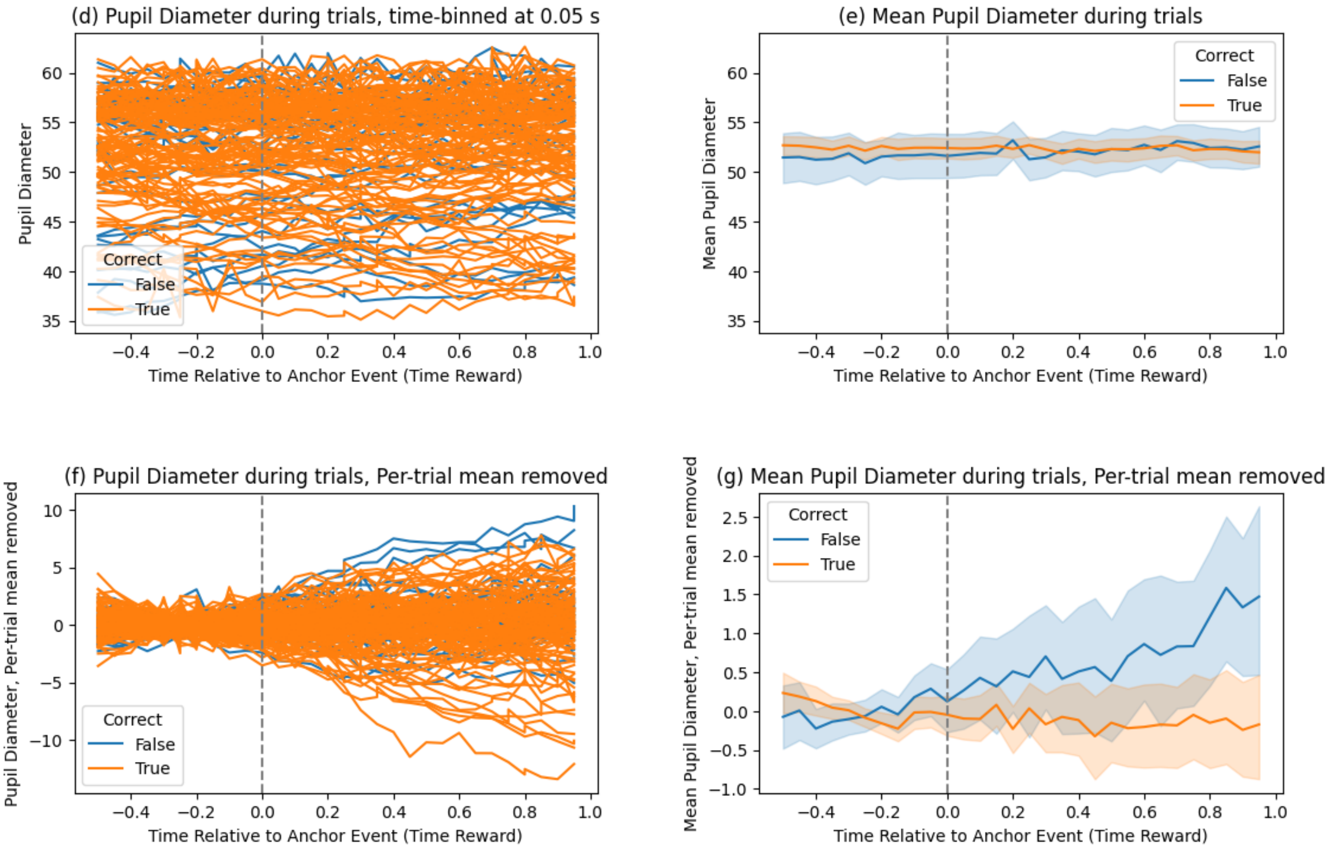

Figure 3: After outlier trials are removed and remaining trials are labeled according to choice correctness, **(d)** samples have their timestamps binned into 50ms buckets. **(e)** The mean trajectory is computed for Correct and Incorrect trials. 95% Confidence intervals are shown. **(f)** Owing to the high trial-specific bias in pupil diameter, each trial has its pre-reward mean pupil diameter computed and subtracted from all trial samples. **(g)** Once per-trial means are removed, mean trajectories are once again computed for Correct and Incorrect trials. 95% Confidence intervals are shown.

`https://www.jneurosci.org/content/30/50/16777`
doi: 10.1523/JNEUROSCI.3967-10.2010