Fixation-aligned Pupillary Response Averaging

Jeff Klingner*
Stanford University

Abstract

We propose a new way of analyzing pupil measurements made in conjunction with eye tracking: *fixation-aligned pupillary response averaging*, in which short windows of continuous pupil measurements are selected based on patterns in eye tracking data, temporally aligned, and averaged together. Such short pupil data epochs can be selected based on fixations on a particular spot or a scan path. The windows of pupil data thus selected are aligned by temporal translation and linear warping to place corresponding parts of the gaze patterns at corresponding times and then averaged together. This approach enables the measurement of quick changes in cognitive load during visual tasks, in which task components occur at unpredictable times but are identifiable via gaze data. We illustrate the method through example analyses of visual search and map reading. We conclude with a discussion of the scope and limitations of this new method.

CR Categories: B.4.2 [Hardware]: Input/Output and Data Communications—Input/Output Devices ; I.4.9 [Computing Methodologies]: Image Processing and Computer Vision—Applications; J.4 [Computer Applications]: Social and Behavioral Sciences—Psychophysiology

Keywords: eye tracking, scan path, fixation, pupillometry, task-evoked pupillary response, fixation-aligned pupillary response

1 Introduction

Scan paths measured by eye trackers provide details about quick-changing attention and cognitive processes that is otherwise unavailable to observation or introspection. This glimpse they provide of internal processes has been applied broadly to investigations of cognitive psychology, perception, and psychophysics. Several research groups in cognitive science and HCI have shown that eye trackers can provide insight into internal mental processes in another way: through cognitive pupillometry (e.g. Pomplun and Sunkara [2003]). When a person experiences increased mental workload (increased vigilance or an increased load on short-term working memory), the resultant brain activity causes a low-level sympathetic dilation of the pupils [Andreassi 2006]. Measuring these workload-related dilations provides a way to study the amount of mental load imposed by various tasks, but such measurements are difficult to conduct because of the dilations' small size.

*e-mail: klingner@stanford.edu

ETRA 2010, Austin, TX, March 22 – 24, 2010.
© 2010 ACM 978-1-60558-994-7/10/0003 $10.00

The magnitude of workload-related dilations is usually less than 0.5 mm, smaller than the magnitude of other simultaneously ongoing pupil changes caused by light reflexes, emotions, and other brain activity, which collectively cause a constant variation in pupil size over a range of a few millimeters. This difference in magnitude between dilations related to cognitive workload and the background pupil variability makes it impossible to distinguish the pupillary response to any one instance of increased cognitive load from the background "noise" of other pupil changes.

One way of addressing this measurement challenge is to record pupil diameter during a task that lasts for many minutes, then either average pupil size over that long period [Pomplun and Sunkara 2003], find consistent short-timescale changes via wavelet transforms [Marshall 2002], or apply frequency-domain analysis [Nakayama and Shimizu 2004; Moloney et al. 2006] to assess aggregate cognitive load during that long task.

An alternative to this aggregation technique allows the measurement of differences in cognitive load at a time scale of fractions of second rather than minutes. This precision is achieved by measuring pupil size during many repetitions of the same short task, then aligning windows of pupil measurements temporally at the moment of task onset and averaging them [Beatty and Lucero-Wagoner 2000]. The averaging operation will preserve any component of the pupil size signal which is correlated in time with the onset of the task (the task-evoked response), while measurement noise and other variation in pupil size not correlated in time with the stimulus will tend to average to zero. As more trials are conducted and included in the average, the ratio achieved between the level of the signal (pupillary responses caused by the task) and the noise (all other pupillary motions) gets bigger, and the time resolution of the average signal improves. Klingner et al. [2008] showed that a remote video eye tracker has sufficient imaging and temporal resolution to allow the use of this signal averaging technique and measured changes in cognitive load during simple arithmetic, memory, and vigilance tasks with a time resolution of 100 ms.

The use of task repetition to measure task-induced dilations forces a tradeoff: fine time resolution is gained, but only for short, simple tasks, in which the cognition of interest occurs at a consistent time soon after an experimenter-controlled stimulus.

In visual tasks such as map reading, chart reading, visual search, and scene comprehension, people shift their attention rapidly and unpredictably, with scan paths being planned on the fly in response to what has been seen so far. It would be useful to measure the dynamics of cognitive load during such tasks, and pupillary response averaging provides good time resolution. But the unpredictability of visual problem solving violates the requirement of signal averaging that the cognitive process being studied happen with predictable timing, which is necessary for aligning the pupillary responses from multiple trials.

This paper describes a new method for assessing cognitive load in such tasks: *fixation-aligned pupillary response aver-*

aging, in which eye fixations are used instead of stimulus or response events to temporally align windows of pupil measurements before averaging. This method enables the detection of quick changes in cognitive load in the midst of long, unstructured tasks, especially visual tasks where fixations on certain points or sequences of points are reliable indicators of the timing of certain task components.

For example, the are certain subtasks that are usually required to read a bar chart, but these subtasks occur at different times from trial to trial and from person to person: e.g. reading the title and axis labels, judging the relative heights of bars, estimating bar alignment with axis ticks, and looking up bar colors in the legend. If we conduct many repetitions of such a chart reading task, changing the details, we can later use scan path analysis to identify all times when somebody compared the height of two bars. We can then align pupil signals from those epochs at the moments of key fixations in the comparison, then average them to determine the average changes in cognitive load during comparison of the height of two bars in a bar chart.

We describe the details of this new averaging method in Section 2, then illustrate its application in two example analyses of cognitive load: one of target discovery in visual search (Section 3.1) and one of references to the legend during map reading (Section 3.2). We conclude with a brief discussion of the method's applicability and limitations.

2 Fixation-aligned pupillary response averaging

Fixation-aligned pupillary response averaging can be broken down into three steps:

1. the identification of subtask *epochs*, short spans of time in which the task component occurs,
2. the temporal alignment of all such subtask epochs, and
3. averaging the aligned epochs.

2.1 Identifying subtask epochs using patterns in gaze data

We use the term *epoch* to refer to short windows of time in which a consistent task component (and therefore a consistent pupillary response) occurs, as well as to the pupil diameter measurements collected during that window of time. Epochs are typically two to ten seconds long. An epoch is characterized by one or more *gaze events*, experimenter-defined fixations or saccades.

Single fixations The simplest gaze event is fixation on a particular spot identified by the experimenter. Epochs defined by single fixations encompass a brief window of time a few seconds long and centered on the fixation. For example, in a visual search task requiring discrimination between targets and distractors, each fixation on a search item determines an epoch containing a visual discrimination subtask. Fixations on targets determine epochs of target recognition (see Section 3.1). In a flight simulation study, each fixation on the altimeter could define an epoch.

Scan paths Gaze events can also be sequences of fixations (scan paths) or saccades from one location to another. For example, fixation on the axis of a bar chart before looking at any of the bars indicates general orientation to the chart, while fixation on the axis immediately after fixating the edge of one of the bars indicates an epoch of axis reading. Comparison of two bars is signaled by several consecutive fixations alternating between them. Epochs defined by scan paths are usually composed of more than one gaze event. For example, in map reading, when somebody looks a symbol in the map, then saccades to the legend to look up the meaning of the symbol, then saccades back to the symbol, these three gaze events comprise a legend reference epoch (see Section 3.2).

In each of these cases, the experimenter defines a sequence of fixations that they expect to reliably occur together with the task component or cognitive process under investigation.

2.2 Aligning pupil data from selected epochs

After all the epochs containing the subtask of interest are identified, they need to be aligned based on the timing of gaze events that make them up.

2.2.1 Temporal Translation

For epochs defined by a single gaze event, like fixation on a particular spot, temporal alignment simply requires translation of each epoch so that their gaze events coincide. Formally, if $P_i(t)$ is pupil diameter as a function of time during epoch i, and g_i is the time of epoch i's gaze event, $T[P_i(t)] = P_i(t + g_i)$ is the temporal translation of $P_i(t)$ which places its gaze event at $t = 0$. Such alignment is done for all epochs that will be averaged. Alignment via temporal translation is illustrated in Figure 1.

2.2.2 Warping

Sometimes, epochs of interest are characterized by multiple gaze events. For example, referencing the legend during a map reading task involves a saccade from the map to the legend, a period of time spent on the legend, and a saccade back to the map. Translation could align the map → legend saccades in all these epochs, but because people don't always dwell on the legend for the same amount of time, the returning legend → map saccades won't line up. If the cognition of interest occurs relative to both points, then signal averaging won't reinforce it.

Porter et al. [2007] faced a similar problem in their analysis of pupil data from tasks of various lengths, in which they needed the start and end times of each task to align. They solved it by setting a fixed task duration and then stretching or compressing the data from each trial to fit in that window.

In the context of task epochs defined by several gaze events, we generalize this method by applying this time-stretching operation to the span of time between each pair of consecutive gaze events. Formally, in an average of n epochs (indexed by i), each of which is defined by m gaze events (indexed by j), the piecewise linear warping of $P_i(t)$ is defined

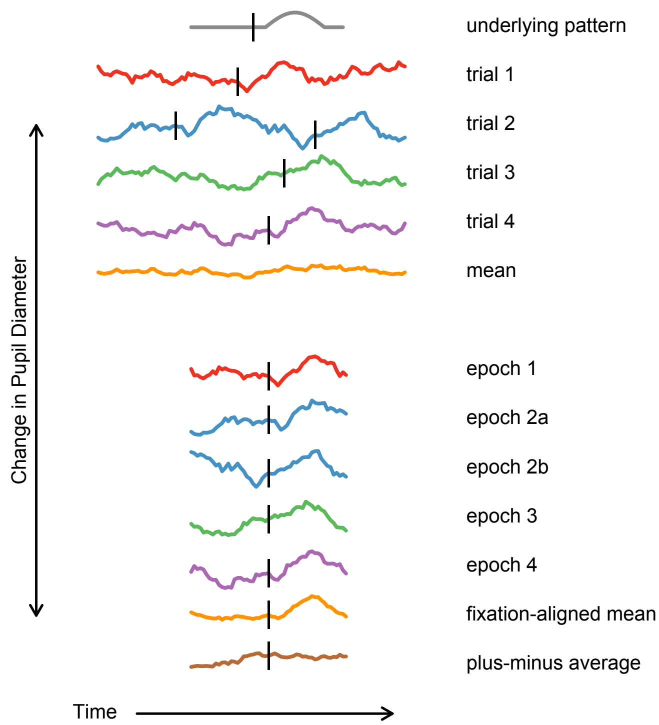

Figure 1: *Illustration of epoch alignment via temporal translation followed by averaging. The top half of the figure shows four simulated trials with gaze events (fixations) occurring at various times. The simulated pupil diameter data for these trials is the sum of random walks (simulating typical background pupil motions) and a dilation response occurring at a fixed delay following the fixation (illustrated at the top of the figure in grey). Because the fixations in these four trials aren't aligned, neither are the pupillary responses, and averaging without translation fails to recover the underlying pattern.*

Epochs aligned by translation are shown in the bottom half of the figure. Because these epochs are aligned on their gaze events, the pupillary responses are aligned too, and averaging the five signals reveals the underlying pupillary response pattern. The final line in the figure is the ±-average of the four aligned signals (see Section 2.3.3), which shows the level of noise present in the mean above it.

In this example, the magnitude of the signal relative to the background pupil noise is exaggerated; in real pupil dilation data, many dozens (sometimes hundreds) of epochs must be averaged before the noise level in the average is low enough to distinguish the pupillary response.

as $W[P_i(t)] =$

$$\begin{array}{ll} P_i\left(g_{i,1} + (g_{i,2} - g_{i,1})\dfrac{t - g_1}{g_2 - g_1}\right) & \text{for} \quad g_1 \le t < g_2 \\ P_i\left(g_{i,2} + (g_{i,3} - g_{i,2})\dfrac{t - g_2}{g_3 - g_2}\right) & \text{for} \quad g_2 \le t < g_3 \\ \vdots & \quad\vdots \\ P_i\left(g_{i,n-1} + (g_{i,n} - g_{i,n-1})\dfrac{t - g_{n-1}}{g_n - g_{n-1}}\right) & \text{for} \quad g_{n-1} \le t \le g_n, \end{array}$$

where $g_{i,j}$ is the time of the jth gaze event in the ith epoch,

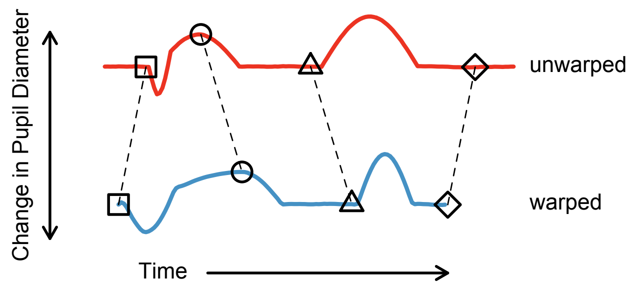

Figure 2: *Illustration of piecewise linear warping applied to a single epoch of pupil diameter data defined by four gaze events. In the original unwarped pupil diameter data, shapes mark the times at which the gaze events occurred. The epoch is divided into segments at the gaze events, and each segment is linearly transformed in time so that the gaze events that bound it are moved into their reference positions in time. These reference positions are determined by averaging the time of occurrence of each gaze event across all epochs (see Section 2.2.2). Figure 3 shows this warping operation applied to several epochs at once before averaging them to reveal pupillary responses that occur with consistent timing with respect to the gaze events.*

and $g_1, g_2, \ldots, g_n$ are the gaze event *reference times*, the mean times of occurrence for each gaze event across all the epochs being aligned: $g_j = \frac{1}{n}\sum_{i=1}^{n} g_{i,j}$. Epoch alignment via piecewise linear warping is illustrated in detail in Figure 2 and applied to averaging several signals in Figure 3. This alignment technique is applied to the analysis of legend references in Section 3.2.

It is important to note that epoch warping is a selective focusing operation. When pupillary responses take place with respect to more than one gaze event, it can reveal them, but at the same time it will obscure any pupillary responses that don't follow that pattern.

2.3 Averaging

2.3.1 Baseline Subtraction

In cognitive pupillometry, the physical quantity of interest is the change in pupil diameter relative to its diameter shortly before the mental activity being studied [Beatty and Lucero-Wagoner 2000]. That is, what matters is dilation (or constriction), not absolute pupil size. The magnitude of dilation responses to simple tasks is independent of baseline pupil diameter and commensurate across multiple labs and experimental procedures [Beatty 1982; Bradshaw 1969; Bradshaw 1970].

In fixation-aligned pupillary response averaging, this means that the pupil data we are averaging needs to be transformed from absolute pupil diameter measurements to dilation measurements. This transformation is accomplished by first determining the baseline pupil size for each epoch by averaging the pupil diameter measurements during the epoch (or during a short window of time at the start of the epoch or surrounding its gaze event), then subtracting that baseline diameter from all pupil diameter measurements made in the epoch. Formally, if the time interval chosen for the baseline extends from $t = b_1$ to $t = b_2$, subtracting the mean pupil diameter during that interval from the full signal gives

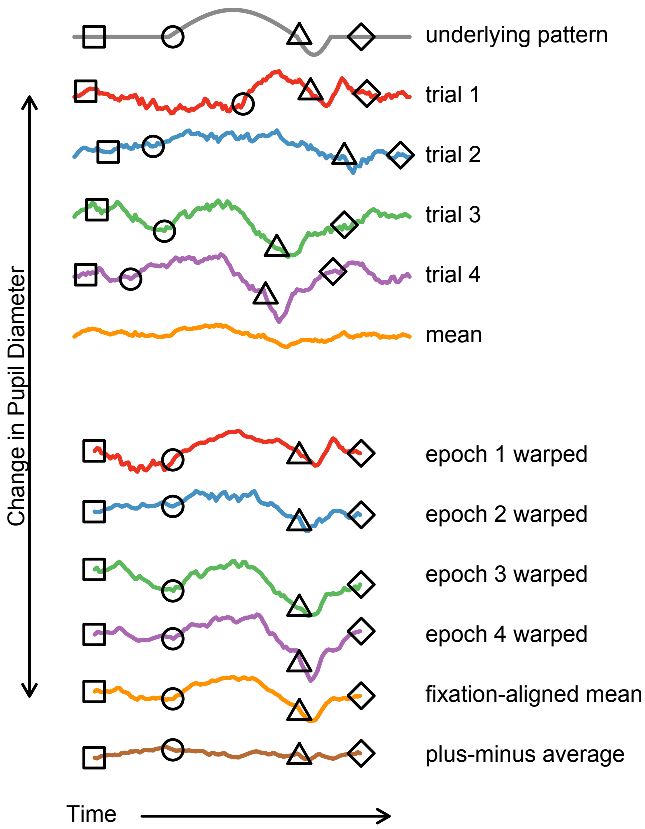

Figure 3: *Illustration of epoch alignment via piecewise linear warping followed by averaging. The top half of the figure shows four simulated trials, each with four gaze events (fixations or saccades) occurring at various times. As in Figure 1, simulated pupil diameter data are the sum of random walks and the indicated pupillary response relative to the four gaze events.*

The bottom half of the figure shows the result of aligning each epoch via piecewise linear warping. The average of the aligned signals reveals the underlying pupillary responses, because they occurred with consistent timing relative to the gaze events.

The final line in the figure is the ±-average of the four warped epochs (see Section 2.3.3), which indicates the level of noise present in the mean above it. As in Figure 1, the magnitude of the signal relative to the background pupil noise is exaggerated.

$B[P_i(t)] = P_i(t) - \frac{\Delta t}{b_2 - b_1} \sum_{t=b_1}^{b_2} P_i(t)$, where Δt is the sampling interval of the eye tracker.

This transformation from diameters to dilations has an important implication for the precision of pupil measurements. For cognitive pupillometry applications, an eye tracker's accuracy in measuring changes in pupil diameter is much more important than its accuracy in measuring absolute pupil size.

2.3.2 Averaging

After all the pupillary epochs have been aligned on their gaze events using translation or warping and their data has been transformed from diameters to dilations via baseline subtraction, the epochs can be averaged using a simple mean: $\overline{P}(t) = \frac{1}{n} \sum_{i=1}^{n} B[T[P_i(t)]]$, or $\frac{1}{n} \sum_{i=1}^{n} B[W[P_i(t)]]$, depending on whether translation or warping is used for alignment. If the data are messy, it may be better to use a trimmed mean or the median instead. The averaged pupillary response $\overline{P}(t)$ is the main object of analysis and is what is graphed in Figures 5, 6, and 8 of the example applications.

Averaging epochs containing consistent pupillary responses preserves the pupillary responses while decreasing the magnitude of the noise in which they are embedded. Because the noise component of the signal is random with respect to the gaze events, the magnitude of the noise average (its standard deviation) decreases in proportion to the square root of the number of epochs included in the average. Cutting the noise by a factor of two requires quadrupling the number of epochs. The actual number of epochs required for a specific experiment depends on the level of measurement noise in the pupillometer and the level of background pupil noise in study participants. In our studies using a remote video eye tracker and tightly controlled visual field brightness (see Section 4.1), we have found that it takes at least 50 epochs to see large pupillary responses (0.2–0.5 mm) cleanly, and hundreds of epochs to reveal pupillary responses smaller than 0.1 mm.

2.3.3 The ±-average

The purpose of averaging aligned pupil dilation data is to preserve the signal of interest (the task-evoked pupillary response) while decreasing the power of signal components not correlated in time with gaze events (the noise). However, the magnitude of the pupillary response being investigated is usually not known *a priori*, so in practice it is difficult to tell whether a feature of the averaged signal $\overline{P}(t)$ is noise or not.

This problem also arises in the analysis of averaged EEG data, where a procedure called the ±-average is used to estimate magnitude of the noise by itself ([Wong and Bickford 1980], originally described as the ±-reference by Schimmel [1967]). Instead of simply adding all the epochs and dividing by n, the epochs are alternately added and subtracted from the running total: $\overline{P}_{\pm}(t) = \frac{1}{n} \sum_{i=1}^{n} P_i(t)(-1)^i$ (only defined for even n). This way, any time-correlated signal will be positive half the time and negative half the time and thus cancel exactly to zero, while any other components of the average, which were already as likely to be positive as negative, will be unchanged and approach zero as a function of n as in the normal average. The average magnitude of $\overline{P}_{\pm}(t)$ is usually a good estimate of the noise power in the standard average. If no pupillary response stands out above this level, then either there is no pupillary response to see, or more trials are required to drive the noise power even lower.

3 Example Applications

While eye trackers have been successfully used for stimulus-locked cognitive pupillometry, it is not obvious that fixation-aligned signal averaging will work. It is possible that the timing of cognitive processes is not consistent enough with respect to eye movements or that the very act of looking around suppresses or interferes with the task-related pupillary response. A successful application of fixation-aligned averaging requires an averaged pupillary response which differs from its corresponding ±-average and any relevant control conditions, and which is consistent with known patterns

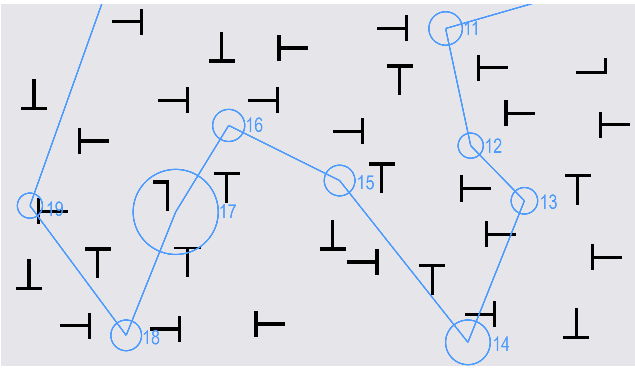

Figure 4: *A fragment of a search field used in our visual search study. Participants searched for L's (targets) in a field of T's (distractors). The scan path from one trial is shown in blue, with circle area proportional to fixation duration and numbers giving the order of fixations. Fixation 17 is a target fixation; all other fixations are non-target fixations.*

of cognition for the studied subtask.

In the following two example analyses, we apply fixation-locked averaging to two well-studied tasks, in order to illustrate its use and to demonstrate its validity. In both tasks, we defined epochs using gaze events we expected to be strongly correlated with shifts in cognitive load, in order to find out whether fixation-aligned averaging revealed the expected shifts in excess of background pupillary noise.

3.1 Visual Search

Visual search has been studied extensively with eye tracking (e.g. Findlay [2009]), and occasionally with pupillometry (e.g. Porter et al. [2007]), though signal averaging has only ever been applied with respect to full task onset and completion. This section summarizes an application of fixation-aligned pupillary response averaging to investigate shifts in cognitive load that occur around the moments of search target discovery.

The averaging used in this example uses the single gaze events and translation alignment described in Section 2.2.1 and illustrated in Figure 1.

3.1.1 Task Description

We designed an exhaustive visual search task in which study participants counted the number of L's (*targets*) in a field of T's (*distractors*) (See Figure 4). The search field contained a variable number of targets, and each trial continued until the participant found them all. Targets were often fixated more than once during a search, with later fixations performed to confirm previously-discovered targets' locations and avoid over-counting.

Equipment We measured participants' gaze direction and pupil size with a Tobii 1750, an off-the-shelf remote infra-red video eye tracker with an integrated display and a single fixed high-resolution bottom-mounted camera, This eye tracker has a sampling rate of 50 Hz and a spatial resolution of 0.5° [Tobii Technologies, Inc. 2007].

Participants We recruited seventeen undergraduate participants, all of whom had normal or corrected-to-normal vision. We excluded people with astigmatism corrections, which can interfere with eye tracking. We compensated participants with Amazon.com gift certificates, the value of which increased with faster and more accurate task performance.

Fixation Identification We segmented scan paths into fixations using the dispersion threshold technique (described by Widdel [1984]; see also Salvucci & Goldberg [2000] for alternatives), with a minimum fixation duration of 160 ms, and a dispersion threshold of 2°.

Consecutive sequences of fixations that all fell within 1.25° of targets were grouped into dwells, within which the fixation that fell closest to the target was labeled as a *target fixation*. Fixations located within the search field but at least 5° from any target, and excluding the first five and last five fixations of the trial, were labeled as *control fixations*, included in the analysis to check for any consistent pupillary response to fixation itself. Both target fixations and control fixations were used as gaze events for selecting pupil data epochs for averaging.

3.1.2 Results

Target fixations vs. control fixations Figure 5 shows the average pupillary response to target and control (non-target) fixations, aligned to the start of the fixation, and showing a few seconds of pre- and post-fixation context. For baseline subtraction, we used a baseline interval 0.4 seconds (20 samples) long, starting 1.75 sec before the fixations. We found a clear difference in pupillary responses to the two different types of event. Fixations far from targets had no consistent pupillary response and so averaged to an approximately flat line, while fixations on targets resulted in a dilation of about 0.06 mm.

Surprisingly, the averaged dilation begins about one second *before* fixation on the target. A further breakdown of the data by difficulty and fixation sequence shows the cause:

Target discoveries vs. target revisits Figure 6 shows the same target fixations, but grouped in two averages, one for all the first fixations on each target (*discoveries*) and one for fixations on targets that have already been fixated during the search (*revisits*). The dilation response to revisits begins at least one second before the target fixation, perhaps reflecting recall of the previously identified target or saccade planning in order to re-confirm its location.

3.1.3 Discussion

As a check on the validity of fixation-aligned pupillary response averaging, this case study was a success. We observed averaged dilations well above the background noise level and which differed substantially between fixations on targets and fixations on non-targets. In addition, the time resolution in the averaged turned out to be fine enough to suggest differences in memory dynamics surrounding target discoveries vs. revisits.

A more thorough analysis of visual search would use more complicated attributes of the scan path, like the fraction of search area that has been covered, to identify additional subtasks, or explore how pupillary responses vary over the

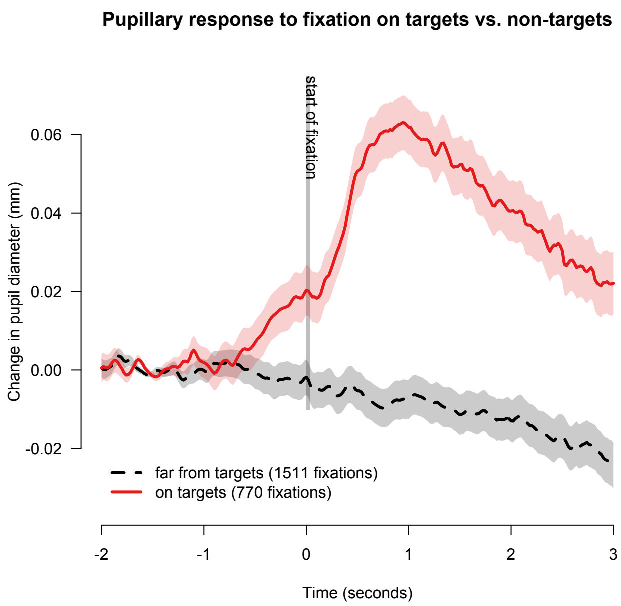

Figure 5: *Fixations on targets vs. fixations on non-targets. Each line in the chart represents the average of many fixations of each type. The shaded regions surrounding each line indicate the standard errors for those averages. Fixations far from targets had no consistent pupillary response and so averaged to an approximately flat line, while fixations on targets resulted in a dilation of about 0.06 mm.*

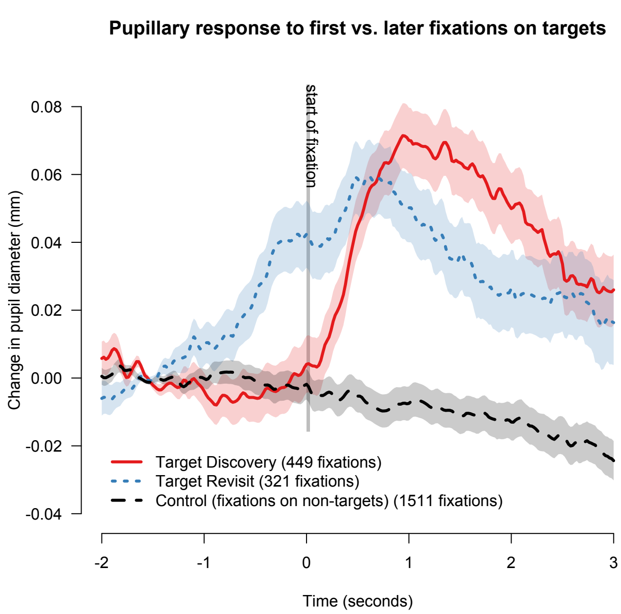

Figure 6: *Average pupillary responses to first (discovery) fixations on targets vs. later (revisit) fixations on targets. The dilation response begins about one second before the fixation for revisits.*

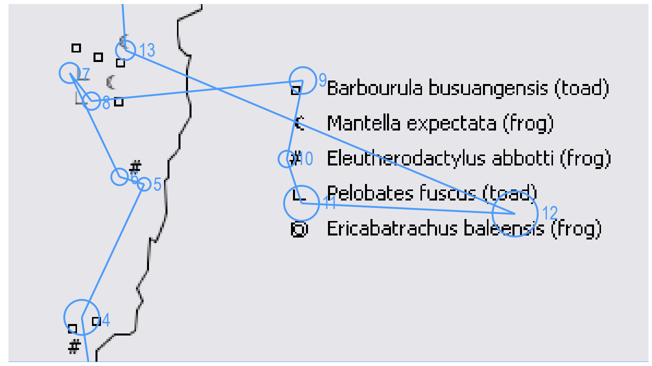

Figure 7: *A fragment of a task map and corresponding legend. A sample scan path is shown in blue, with circle area proportional to fixation duration and numbers indicating the order of fixations. A legend reference epoch begins at the time of fixation 7 and ends at the time of fixation 13. The gaze events used for alignment of this epoch are the 8 → 9 and 12 → 13 saccades.*

course of the search. Additionally, the differences in timing and magnitude of pupillary reactions could be analyzed between subjects, or with respect to task performance.

3.2 Map Legend Reference

The second example application of fixation-aligned pupillary response averaging uses more complicated epochs, defined using multiple gaze events and aligned with warping.

3.2.1 Task Description

In a study of map reading, participants examined a fictitious map showing the locations of frog and toad habitats. The map uses abstract symbols to show places where frogs and toads live, with each symbol standing for a different species. The symbols are identified in a legend providing the species name and classification as frog or toad (Figure 7). In reading the map, participants must look up the abstract symbols in the legend to learn which of them correspond to frogs and which correspond to toads. It is these legend references which we analyze here.

Participants & Apparatus This study included fifteen undergraduates, distinct from those in the first study but subject to the same selection criteria. The eye tracker was the same.

3.2.2 Identifying legend reference epochs

Epochs of pupil measurements encompassing legend references were identified using scan paths. We defined a legend reference as fixation on a cluster of map symbols, followed by a fixation on the legend, followed by a return saccade to that same cluster of map symbols. An example epoch is shown in Figure 7. We used the saccades to and from the legend as gaze events on which to align epochs via piecewise warping before averaging (see Section 2.2.2). Baseline intervals used for baseline pupil diameter subtraction were 0.4 sec (20 samples) long, starting 1.5 sec before the saccade from the map to the legend. We calculated fixations using dispersion threshold clustering as described in Section 3.1.1.

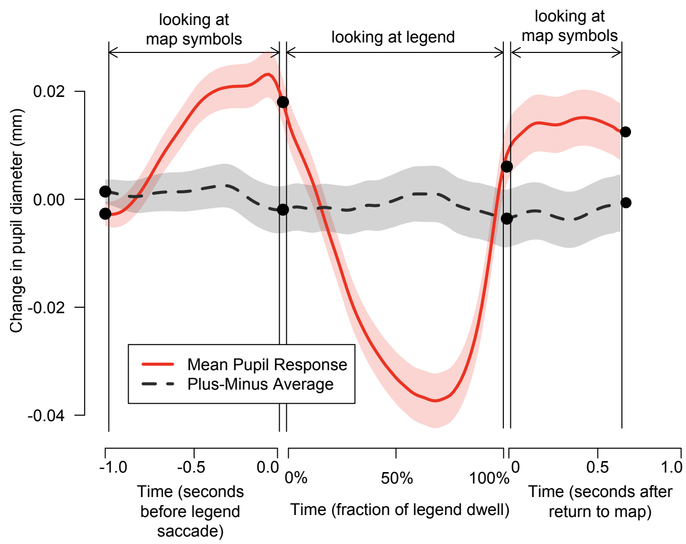

Figure 8: *Average pupillary response to 925 legend references in a map reading task. Black circles indicate reference gaze event times. The semi-transparent regions bounding each curve show the standard errors of the mean at each time for the plotted average. The average dwell on the legend lasted about 1.2 seconds. The average pupillary response includes a 0.02 mm dilation prior to saccade to the legend, a 0.06 mm constriction and recovery while looking at the legend, and a return to the map at a slightly higher pupil diameter.*

3.2.3 Results

We expected a momentary dilation preceding each legend reference, caused by the need to store the symbol or symbols being looked up in visual working memory during the saccade to the legend and the search there for the matching symbol(s). This pattern did emerge in the averaged pupil response, along with other changes also correlated with the legend reference saccades (Figure 8). On average, participants' pupils contracted while looking at the legend before recovering nearly to their pre-saccade diameter. Although the changes were small (on the order of 0.05 mm), we collected enough epochs (925 legend references) that these changes stood out above the noise level indicated by the ±-average.

3.2.4 Discussion

Like the visual search study, this application of fixation-aligned pupillary response averaging succeeded. We found a pupillary response evoked by the subtask of referencing the map legend which substantially differed from its corresponding ±-average, though in this case several hundred trials were required to reveal the response.

The changes in pupil diameter which we observed are intriguing; in addition to the simple pre-reference dilation we expected, we also observed a pupil constriction during the dwell on the legend. This pattern is consistent with the loading of visual working memory with the symbol to be looked up, the release of that memory once the symbol has been located in the legend, and a final increase in load as the participant's running count of frogs is updated depending on the symbol's classification.

4 Conclusions

This paper describes fixation-aligned pupillary response averaging, a new method for combining synchronized measurements of gaze direction and pupil size in order to assess short-term changes in cognitive load during unstructured visual tasks. Components of the visual tasks with consistent demands but variable timing are located by analyzing scan paths. Pupil measurements made during many instances of each task component can then be aligned in time with respect to fixations and averaged, revealing any consistent pupillary response to that task component.

This new mode of analysis expands the scope of tasks that can be studied using cognitive pupillometry. With existing stimulus-locked averaging methods, only shifts in cognitive load that occur relative to experimenter-controlled stimuli (e.g. Klingner et. al [2008]) are measurable, but with fixation-aligned averaging, pupillary responses can also be used to study any shifts in cognitive load that occur consistently with respect to patterns of attention detectable in gaze direction data.

In the example study of visual search described in Section 3.1, the timing differences in pupillary responses to target discoveries and revisits, which show the recall of previously-visited targets, are only detectable through fixation-aligned averaging. Similarly, the shifts in cognitive load surrounding subject-initiated legend references described in Section 3.2 could only be detected by determining the timing of those legend references using gaze direction data and then using that timing information to align and average pupil diameter measurements.

There are many other tasks that could be studied using this method. Reading, for example, which has been studied using eye tracking [Rayner 1998] and pupillometry [Just and Carpenter 1993] separately, has many gaze-signaled cognitive events, such as back-tracking, which could be studied with fixation aligned averaging of pupil measurements.

4.1 Limitations

Simple, short task components Fixation-aligned averaging can be employed when subtasks epochs can be reliably identified from gaze direction data and when changes in cognitive load during these epochs occur with consistent timing relative to the gaze events that define the epochs. In practice, this limits both the specificity and duration of subtasks that can be studied. For specificity, either the general task must be constrained enough or the gaze events defined specifically enough that cognitive processes are consistent across epochs. For example, in our map-reading task, we only considered legend references which began and ended at the same point in the map, to avoid including fixations on the legend that were simple orientation to the display, or otherwise did not involve looking up a particular symbol.

Even when epochs containing consistent cognitive processes are identified, the requirement that the timing of those processes be consistent with respect to the fixations and saccades which define the epochs is generally only satisfied for short epochs, in practice usually 2–5 seconds long.

Restrictions on task display Because pupils respond reflexively to the brightness [Loewenfeld 1999] and contrast level [Ukai 1985] of whatever is currently fixated, the task display must be designed with spatially uniform brightness and contrast. In addition, epochs occurring soon after changes to the display can be contaminated by reflex dilations to motion cues [Sahraie and Barbur 1997].

In both the studies presented here, we used static, low-contrast stimuli with spatially-uniform brightness (Figures 4 and 7). In addition, we left a large margin between the edge of the stimulus and the boundary of the display, because when subjects fixate near that boundary, their field of view includes the display bezel and the wall behind, both of which are difficult to match in brightness to the stimulus itself.

Restrictions on interaction Any user interaction, such as mouse or keyboard use, needs to be well separated in time from the subtask epochs studied, because the preparation and execution of motor actions also causes momentary pupil dilations [Richer and Beatty 1985]. We designed both our visual search and map reading studies without any interaction during each task and had to exclude any subtask epochs that occurred within three seconds of the button-pushes that participants used at the end of each trial to signal their completion.

Acknowledgements

This work was funded by the Stanford Regional Visual Analytics Center, through the U.S. Department of Energy's Pacific Northwest National Laboratory. Our eye tracker was funded by the Stanford MediaX project and the Stanford School of Engineering.

References

ANDREASSI, J. L. 2006. *Pupillary Response and Behavior*, 5th ed. Routledge, 289–307.

BEATTY, J., AND LUCERO-WAGONER, B. 2000. The pupillary system. In *Handbook of Psychophysiology*, J. T. Cacioppo, L. G. Tassinary, and G. Berntson, Eds., 2nd ed. Cambridge University Press, 142–162.

BEATTY, J. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin 91*, 2, 276–292.

BRADSHAW, J. L. 1969. Background light intensity and the pupillary response in a reaction time task. *Psychonomic Science 14*, 6, 271–272.

BRADSHAW, J. L. 1970. Pupil size and drug state in a reaction time task. *Psychonomic Science 18*, 2, 112–113.

FINDLAY, J., AND SQUIRE, L. R. 2009. Saccades and visual search. In *Encyclopedia of Neuroscience*. Academic Press, Oxford, 429–436.

JUST, M. A., AND CARPENTER, P. A. 1993. The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology 47*, 2, 310–339.

KLINGNER, J., KUMAR, R., AND HANRAHAN, P. 2008. Measuring the task-evoked pupillary response with a remote eye tracker. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, ACM, Savannah, Georgia, 69–72.

LOEWENFELD, I. 1999. *The Pupil: Anatomy, Physiology, and Clinical Applications*, 2nd ed., vol. 1. Butterworth-Heinemann, Oxford, UK.

MARSHALL, S. 2002. The index of cognitive activity: measuring cognitive workload. In *Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants*, 7–5–7–9.

MOLONEY, K. P., JACKO, J. A., VIDAKOVIC, B., SAINFORT, F., LEONARD, V. K., AND SHI, B. 2006. Leveraging data complexity: Pupillary behavior of older adults with visual impairment during HCI. *ACM Transactions on Computer-Human Interaction 13*, 3, 376–402.

NAKAYAMA, M., AND SHIMIZU, Y. 2004. Frequency analysis of task evoked pupillary response and eye-movement. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, ACM, San Antonio, Texas, 71–76.

POMPLUN, M., AND SUNKARA, S. 2003. Pupil dilation as an indicator of cognitive workload in Human-Computer interaction. In *Proceedings of the International Conference on HCI*.

PORTER, G., TROSCIANKO, T., AND GILCHRIST, I. D. 2007. Effort during visual search and counting: Insights from pupillometry. *The Quarterly Journal of Experimental Psychology 60*, 2, 211.

RAYNER, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin 124*, 3, 372–422. PMID: 9849112.

RICHER, F., AND BEATTY, J. 1985. Pupillary dilations in movement preparation and execution. *Psychophysiology 22*, 2, 204–207.

SAHRAIE, A., AND BARBUR, J. L. 1997. Pupil response triggered by the onset of coherent motion. *Graefe's Archive for Clinical and Experimental Ophthalmology 235*, 8, 494–500.

SALVUCCI, D. D., AND GOLDBERG, J. H. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ACM, New York, NY, 71–78.

SCHIMMEL, H. 1967. The (±) reference: Accuracy of estimated mean components in average response studies. *Science 157*, 3784 (July), 92–94.

TOBII TECHNOLOGIES, INC., 2007. Tobii 1750.

UKAI, K. 1985. Spatial pattern as a stimulus to the pupillary system. *Journal of the Optical Society of America A 2*, 7, 1094–1100.

WIDDEL, H. 1984. Operational problems in analysing eye movements. In *Theoretical and Applied Aspects of Eye Movement Research*, A. G. Gale and F. Johnson, Eds. Elsevier, New York, 21–29.

WONG, P. K., AND BICKFORD, R. G. 1980. Brain stem auditory evoked potentials: the use of noise estimate. *Electroencephalography and Clinical Neurophysiology 50*, 1-2, 25–34. PMID: 6159189.