Beyond Gaze: Preliminary Analysis of Pupil Dilation and Blink Rates in an fMRI Study of Program Comprehension

Norman Peitek
Leibniz Institute for Neurobiology
Magdeburg, Germany

Janet Siegmund
University of Passau
Passau, Germany

Chris Parnin
NC State University
Raleigh, North Carolina, USA

Sven Apel
University of Passau
Passau, Germany

André Brechmann
Leibniz Institute for Neurobiology
Magdeburg, Germany

ABSTRACT

Researchers have been employing psycho-physiological measures to better understand program comprehension, for example simultaneous fMRI and eye tracking to validate top-down comprehension models. In this paper, we argue that there is additional value in eye-tracking data beyond eye gaze: Pupil dilation and blink rates may offer insights into programmers' cognitive load. However, the fMRI environment may influence pupil dilation and blink rates, which would diminish their informative value. We conducted a preliminary analysis of pupil dilation and blink rates of an fMRI experiment with 22 student participants. We conclude from our preliminary analysis that the correction for our fMRI environment is challenging, but possible, such that we can use pupil dilation and blink rates to more reliably observe program comprehension.

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**; **Empirical studies in HCI**;

KEYWORDS

program comprehension, functional magnetic resonance imaging, eye tracking, pupil dilation, blink rates

ACM Reference Format:
Norman Peitek, Janet Siegmund, Chris Parnin, Sven Apel, and André Brechmann. 2018. Beyond Gaze: Preliminary Analysis of Pupil Dilation and Blink Rates in an fMRI Study of Program Comprehension. In *EMIS '18: Symposium on Eye Movements in Programming, June 14–17, 2018, Warsaw, Poland*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3216723.3216726

1 INTRODUCTION

Program comprehension is an internal, difficult to measure cognitive process. Hence, researchers have been moving to observing program comprehension with novel physiological measures, for example, eye tracking [Bednarik and Tukiainen 2006; Busjahn et al. 2015], functional magnetic resonance imaging (fMRI) [Floyd et al. 2017; Siegmund et al. 2014, 2017], or electroencephalography (EEG) [Fritz et al. 2014; Lee et al. 2017]. Numerous studies moved toward multi-modality, particularly combining eye tracking and measuring cognitive load. For example, Fritz et al. used EEG, eye tracking, and electrodermal activity to predict task difficulty [Fritz et al. 2014]. Fakhoury et al. used functional near-infrared spectroscopy (fNIRS) and eye tracking to show how the quality of identifier names affects cognitive load [Fakhoury et al. 2018]. We previously proposed the simultaneous observation of program comprehension with fMRI and eye tracking to understand the effect of rich identifier names on top-down comprehension [Peitek et al. 2017]. In sum, eye tracking appears to be a valuable measure of programmers' visuo-spatial attention: either on its own or in addition to a cognitive-load measure. However, many studies only considered the eye gaze, that is, the fixations (locations, lengths) and saccades (speeds, lengths). Secondary data, such as pupil dilation or blink rates, do not appear to be commonly investigated. This is surprising, as cognitive science has accepted pupil dilation (cf. Section 2.1) and blink rates (cf. Section 2.2) as valuable measures to observe participants' mental states [Eckstein et al. 2017].

In this paper, we explore whether pupil dilation and blink rates offer insights in studying program comprehension. To this end, we conducted a pilot fMRI study with 22 student participants, which included simultaneous eye tracking. Our long-term goal of observing pupil dilation and blink rates in addition to brain activation via fMRI is to detect cognitive events of smaller granularity. While fMRI allows us to observe programmers' cognitive load on a larger scope (e.g., difficulty to comprehend a class), observing the effect of comprehending individual lines may currently be impossible with fMRI. The temporal resolution is 1 to 2 seconds, which means that we may miss short-lived cognitive events, such as a programmer stumbling over an unexpected implementation of a single line. We hope that the integration of pupil dilation lets us detect exact lines that cause programmers to struggle. Furthermore, pupil dilation and blink rates may offer additional measures to observe cognitive load and, as such, can help us to explain some of our fMRI results.

Despite the well-defined conditions in fMRI, with the participant's head in a fixed position and an overall constant illumination inside the magnet bore, we still have to deal with an inconsistent display brightness due to different snippet lengths. In this paper, we take a first look into whether we can correct for the changing display brightness and whether pupil dilation and blink rates are a promising measure for our studies on program comprehension.

EMIS '18, June 14–17, 2018, Warsaw, Poland

ACM ISBN 978-1-4503-5792-0/18/06…$15.00
https://doi.org/10.1145/3216723.3216726

2 BACKGROUND

Researchers have been employing a variety of measures to observe program comprehension. Every measure offers a unique perspective and thus different insights. For example, fMRI allows us to observe brain activation and thus identify neural correlates of cognitive processes during program comprehension [Gazzaniga et al. 2013]. Eye tracking offers insight into the visuo-spatial attention of programmers [Holmqvist et al. 2011]. Pupil dilation and blink rates are two measures that have been used successfully in other fields (see below), but as far as we are aware of, they are rarely applied in studies of program comprehension [Obaidellah et al. 2018].

2.1 Pupil Dilation

One measurable property of eye tracking is pupil dilation, which is a task-evoked pupillary response. Pupil dilation has been shown to reflect cognitive load and has been used as a measure when investigating tasks involving working memory, reasoning, or reading [Beatty and Lucero-Wagoner 2000]. For example, early work of Beatty and Kahneman showed that an increase in the number of digits to be remembered correlates positively with pupil dilation [Beatty and Kahneman 1966]. Similarly, Hess and Polt demonstrated that pupil dilation correlates with the difficulty of mathematical calculations [Hess and Polt 1964]. Later work by Just and Carpenter indicated that pupil dilation generally correlates with the amount of mental processing [Just and Carpenter 1993]. Although mental fatigue does not change the pupil-dilation baseline, it may slow down the pupillary response [Hopstaken et al. 2015]. Overall, research has established that pupil dilation is an accurate measure of mental states [Hartmann and Fischer 2014; Laeng et al. 2012].

Some researchers have suggested to use pupil dilation as a measure in program-comprehension research. Nolan et al. proposed a study for novice programmers, which are learning to program and are observed with remote eye tracking [Nolan et al. 2015]. The goal was to remotely measure cognitive load. Similarly, Ford et al. proposed to identify mental states based on eye-tracking data (i.e., pupil dilation, saccades, blink rates) during remote interviews, which would guide interviewers only interrupting during light thinking phases [Ford et al. 2015]. However, there appears to be no report on a completed study. Thus, the work reported in this paper will be the first to analyze pupil dilation as a measure to observe cognitive load during program comprehension.

2.2 Blink Rates

Blink rates correlate with "levels of dopamine in the central nervous system, and can reveal processes underlying learning and goal-directed behavior" [Eckstein et al. 2017]. Blink rates are determined at two levels: the resting baseline and task-evoked blink rate. A higher individual's blink-rate baseline is related with "better cognitive flexibility but worse maintenance" [Eckstein et al. 2017]. Some research suggests that "higher blink rate at baseline is related with lower distractibility on tasks that place high demands on working memory" [Eckstein et al. 2017]. How this affects blink rates in programming is an interesting research question.

Another factor is that blink rates increase with fatigue [Stern et al. 1994]. While we interleave our task conditions, we have to consider this effect when comparing individual tasks (e.g., a snippet shown at the end of an experiment might cause more blinks as participants are more tired). Also environmental factors, such as air humidity or room temperature, can also influence the blink rate [Doughty 2001]. However, these are controlled in the fMRI environment.

Ford et al. appears to be the only one suggesting that blink rates may be an interesting measure in their proposal for observing technical remote interviews [Ford et al. 2015]. In follow-up work by Behroozi et al., blink rates were not a significant measure to distinguish mental states. However, blink *durations* showed significant differences [Behroozi et al. 2018]. As blink duration reliably increases with cognitive workload, we include it as a measure into our preliminary analysis [Benedetto et al. 2011; Chen et al. 2011].

3 EXPERIMENT

The literature from psychology and cognitive science suggests that pupil dilation and blink rates may be valuable measures in future studies of program comprehension. Thus, we conducted a pilot study to obtain a first evaluation of the two measures by observing program comprehension with simultaneous fMRI and eye tracking with 22 student participants. To this end, we replicated our previous fMRI study on contrasting bottom-up and top-down comprehension [Siegmund et al. 2017], which included five different conditions. We manipulated the richness of identifier names (i.e., *beacons* [Brooks 1983]) and familiarized participants with a subset of snippets during a training session. We used tasks of finding syntax errors as a control condition, resulting in the following trial:

- Top-down comprehension [Trained, including beacons]
- Bottom-up comprehension
- Top-down comprehension [Trained, no beacons]
- Top-down comprehension [Untrained, including beacons]
- Finding syntax errors

In addition to the programming tasks, we included a d2 attention task[1] and rest condition to provide a baseline for fMRI analysis.

The study was conducted on a 3-Tesla fMRI scanner,[2] equipped with a 32-channel head coil at the Leibniz Institute for Neurobiology in Magdeburg, Germany. We used an MRI-compatible EyeLink 1000[3] eye tracker for simultaneous measurement of eye movements. The EyeLink eye-tracker offers 1000 Hz temporal resolution, <0.5° average accuracy, and 0.01° root mean square (RMS). The eye-tracker collected eye gazes, events (i.e., fixations, blinks, saccades), and pupil dilation. Participant demographics, a replication package, and details on our methods are available on the project Web site.[4]

Eye tracking in the fMRI scanner is challenging due to the difficult angle of the eye tracker to a participant's eyes. In our study, only 8 out of 22 participants resulted in stable and precise eye-tracking data that unequivocally locates the eye gaze at the level of single lines, that is with a constant precision of <1°. However, the current pilot data are sufficient to use pupil dilation and blink rates when manually analyzing eye gaze direction. We will focus on participants with stable and precise eye tracking in this exploratory analysis. Our goal for the analysis is to evaluate whether we can

[1] d2 is a test of attention in which participants scan through a row of letters and decide for each letter whether it is a *d* with two marks [Brickenkamp et al. 2010]
[2] Philips Achieva dStream, Best, The Netherlands
[3] SR Research Ltd, Ottawa, Ontario, Canada, http://www.sr-research.com
[4] https://github.com/brains-on-code/simultaneous-fmri-and-eyetracking

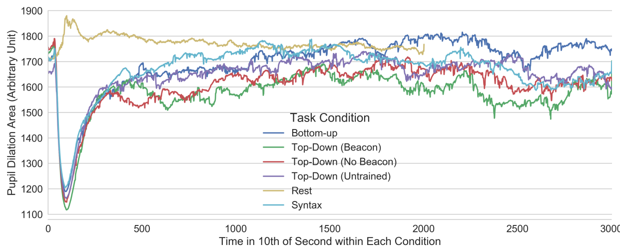

Figure 1: Pupil Dilation over Time for Each Condition

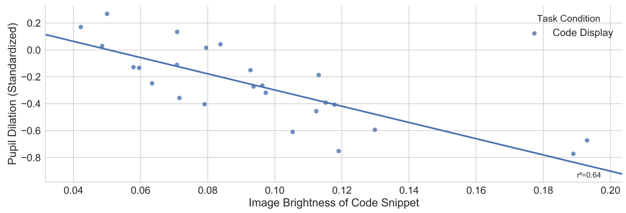

Figure 2: Snippet Brightness and Pupil Dilation

consider pupil dilation and blink rates as a measure in future experiments. Can we find the expected correlation between cognitive load and pupil dilation?

4 PRELIMINARY DATA ANALYSIS

4.1 Pupil Dilation

For a comparison across participants, we normalized the pupil-dilation data for each participant to zero mean and unit standard deviation (z-score).

Screen-Brightness Correction. Any analysis of pupil-dilation data assumes that a change in pupil dilation is only caused by a transition in mental state and not by the environment (e.g., ambient light). The environment around the fMRI scanner ensures that the ambient light does not change. However, we present code snippets to the participants via a small plastic screen on which the stimuli are projected. To reduce eye strain, we use white text on a black background. As our snippets are not uniform in length, each snippet is different in their perceived brightness. Thus, the snippets may influence pupil dilation by their brightness, independent of the change in cognitive load. This effect is visible in Figure 1, where the adjustment from a dark rest condition to a brighter comprehension condition is apparent: The pupil dilation briefly drops at the beginning of the comprehension conditions as the pupils responds to the brighter light.

Because a standardized brightness of the stimuli is infeasible for us (i.e., snippets will generally differ in length), we will need to correct the baseline pupil dilation for each snippet. Figure 2 shows that there is a general trend to a lower pupil dilation with brighter snippets. We computed each snippet's brightness variable as relative luminance of the RGB color space for each image (as an average of each pixel) [Anderson et al. 1996].

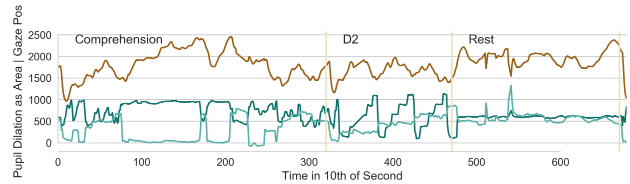

Figure 3: First Minute of Participant 1. Pupil Dilation, Gaze Position on x-Axis and y-Axis

The $r^2 = 0.64$ value shows that the screen brightness largely explains the difference in baseline pupil dilation. However, there still is 36% of variance that is not explained by screen brightness, which could include factors such as comprehension strategy (top-down versus bottom-up), individual difficulty to comprehend a snippet, or general error.

Eye-Movement Correction. So far, we have considered the average pupil dilation throughout a task. To extract the maximum value from pupil dilation, we would like to evaluate changes of pupil dilation within a task. Brisson et al. have shown that pupil size is overestimated for rightward and upward gaze and underestimated for leftward and downward gaze for our used EyeLink 1000 eye-tracker [Brisson et al. 2013], and they recommend to proceed with methodological caution. Ideally, tasks are used that do not require any eye movement. However, this is infeasible, as program comprehension requires programmers to quickly move across the screen. Thus, long and fast saccades are necessary and common. The influence of eye movements on the pupil dilation questions how much we can trust the data, even if corrected for screen brightness.

In Figure 3, we show a pupil-dilation chart of an individual participant across a few tasks. A few significant spikes are noticeable (especially at the rest condition). The eye tracker's manual warns that a fast change in pupil angle can lead to a flawed measurement of pupil dilation. Based on the data, we confirm that fast saccades impair the pupil-dilation accuracy. In Figure 3, we show that the pupil dilation spikes appear to correlate with sudden eye movements, in particular, on the vertical axis (e.g., at time 520 and 550). Movements on the x-axis, that is, leftward and rightward gazes, also seem to have an influence on our pupil-dilation data. For example, the pupil dilation in Figure 3 for the d2 attention tasks appears to rhythmically move with the eye gaze.

For future analysis, we would like correct the pupil dilation for such eye movements. Currently, we evaluate the correction algorithm from Brisson et al. [Brisson et al. 2013] to further improve the accuracy of the pupil-dilation data. As we only consider average pupil dilation for the following sections, we will not apply the correction algorithm.

Task Condition. In our previous fMRI study, we have shown that top-down comprehension is easier for programmers [Siegmund et al. 2017]. Can we support this finding based on pupil dilation? In Table 1, we show the average pupil dilation across all participants and tasks per condition. Top-down comprehension with or without a beacon reveals a difference in pupil dilation, even though the screen brightness is almost identical. This may hint at participants using a beacon for an eased comprehension of a snippet.

Table 1: Mean Perceived Brightness, Pupil Dilation (z-score), Blink Rate (Count/Minute), and Blink Duration (ms) for Each Condition.

Condition	Perceived Brightness	Pupil Dilation	Blink Rate	Blink Duration
Top-Down (Beacon)	3.01	-0.49	12.0	245
Top-Down (No Bea.)	3.07	-0.41	11.6	341
Top-Down (Untr.)	2.09	-0.25	7.6	346
Bottom-Up	2.13	-0.07	8.4	340
Syntax	1.89	-0.11	10.4	322
d2	2.15	0.16	10.1	327
Rest	0.01	0.31	7.9	490

Bottom-up comprehension shows a slightly higher pupil dilation than untrained top-down comprehension, which may be due to the additional mental effort necessary to comprehend a snippet. The d2 attention task has a much higher pupil dilation, even though the brightness is similar to the bottom-up snippets. None of these findings are statistically significant as there were only five trials for each condition. As the other conditions are not balanced in brightness, and we do not have the capabilities to counteract the brightness yet, we will not compare them.

Snippet Complexity. Next, we correlated snippet complexity, measured in DepDegree [Beyer and Fararooy 2010] and Halstead complexity [Halstead 1977], with the observed pupil dilation. The results indicate that, with increasing snippet complexity, the pupil dilation decreases, but only explains a part of the variance ($r^2 = 0.22$, $r^2 = 0.38$, respectively). We also have to consider that snippet complexity is generally correlated with snippet brightness ($r = 0.56$), as longer snippets tend to be more complex. It appears that snippet complexity actually increases pupil dilation (when controlling for screen brightness), a result that is supported by the literature.

Task Difficulty. Instead of considering the objective snippet complexity, we may observe a more significant difference in subjective difficulty. To this end, we correlated the averaged pupil dilation with the task correctness and response time. The results show that there is no correlation ($r^2 = 0.00$, $r^2 = 0.01$, respectively).

Stepwise Regression. Finally, we conducted a stepwise regression to select the significant variables that explain the variance in pupil dilation. The result revealed that screen brightness ($p < 0.001$) and response time ($p < 0.1$) significantly influence pupil dilation.

4.2 Blink Rates and Durations

Next, we analyzed the blink rates and blink duration. The blink rate is the number of blinks that occur over time. It is usually expressed as number of blinks per minute [Holmqvist et al. 2011]. We measured the blink duration in milliseconds.

Our experiment in the fMRI scanner lasted for around 28 minutes. There was no notable increase in blink rates towards the end of the experiment. Fatigue does not seem to have set in after 28 minutes yet, and blinks are not influenced by eye movements, thus we can analyze the blink rates without correction.

Task Condition. In Table 1, we show the average blink rate and blink duration for each task condition. We cannot identify any clear-cut patterns. However, it is notable that top-down comprehension with a beacon has a shorter average blink duration than any other comprehension task.

Snippet Complexity. We correlated the observed blink rate and duration with snippet complexity, again measured in terms of DepDegree and Halstead complexity. The results indicate that, with increased complexity, the blink rate decreases, but the effect explains only a minor part of the variance ($r^2 = 0.15$, $r^2 = 0.07$, respectively). Blink duration has the same negative correlation, but explains less variance ($r^2 = 0.09$, $r^2 = 0.04$, respectively)

Task Difficulty. We also correlated the observed blink rate with the task correctness and response time. We found a weak negative correlation between blink rate and task correctness ($r^2 = 0.01$, $r^2 = 0.12$, respectively). The same holds for blink duration ($r^2 = 0.05$, $r^2 = 0.00$, respectively).

Future Analysis. This first look at blink rates and durations does not reveal any meaningful insights. Nevertheless, we will conduct further detailed analyses, which likely reveal some insights. For example, we plan to consider the point in time of each blink. Did more or longer blinks occur after participants solved the task? Did blinks occur particularly often at a specific time? This way, we hope to discover parts of a code that are particularly difficult.

5 CONCLUSION

We set out to include pupil dilation and blink rates as promising measures to increase the reliability of measuring program comprehension with fMRI. However, our initial study and analysis showed that, so far, pupil dilation largely depends on screen brightness, and that blink rates and blink durations do not seem to follow a pattern. Nevertheless, our data indicate that, although screen brightness is identical, the pupil dilation varies (e.g., depending on the task condition), which suggests that there is an effect of cognitive load on pupil dilation. Furthermore, due to the strong linear relationship between screen brightness and pupil dilation, we may be able to correct for the changing brightness in a thorough analysis.

Our preliminary analysis confirms our belief that there could be value beyond the usual gaze analysis of eye tracking. Pupil dilation and blink rates may provide interesting insight for studies of program comprehension. However, before we can use them, we need to find ways to control for environmental factors, such as screen brightness, and develop appropriate analyses.

ACKNOWLEDGMENTS

We thank all participants of the fMRI study. Furthermore, we thank Andreas Fügner, Anke Michalsky, and Jörg Stadler for their technical support during pilot studies and fMRI data acquisition.

Siegmund's and Brechmann's work is supported by DFG grants SI 2045/2-1 and BR 2267/7-1. Siegmund's work is further funded by the Bavarian State Ministry of Education, Science and the Arts in the framework of the Centre Digitisation.Bavaria (ZD.B). Parnin's work is supported by the National Science Foundation under grant number 1755762. Apel's work has been supported by the German Research Foundation (AP 206/6).

REFERENCES

Matthew Anderson, Ricardo Motta, Srinivasan Chandrasekar, and Michael Stokes. 1996. Proposal for a Standard Default Color Space for the Internet—rgb. In *Color and Imaging Conference*, Vol. 1996. Society for Imaging Science and Technology, 238–245.

Jackson Beatty and Daniel Kahneman. 1966. Pupillary Changes in Two Memory Tasks. *Psychonomic Science* 5, 10 (1966), 371–372.

Jackson Beatty and Brennis Lucero-Wagoner. 2000. The Pupillary System, Handbook of Psychophysiology, Cacioppo, Tassinary & Berntson. (2000).

Roman Bednarik and Markku Tukiainen. 2006. An Eye-Tracking Methodology for Characterizing Program Comprehension Processes. In *Proc. Symposium on Eye Tracking Research & Applications (ETRA)*. ACM, 125–132.

Mahnaz Behroozi, Alison Lui, Ian Moore, Denae Ford, and Chris Parnin. 2018. Dazed: Measuring the Cognitive Load of Solving Technical Interview Problems at the Whiteboard. In *Proc. Int'l Conf. Software Engineering (ICSE)*. IEEE, 4 pages.

Simone Benedetto, Marco Pedrotti, Luca Minin, Thierry Baccino, Alessandra Re, and Roberto Montanari. 2011. Driver Workload and Eye Blink Duration. *Transportation Research Part F: Traffic Psychology and Behaviour* 14, 3 (2011), 199–208.

Dirk Beyer and Ashgan Fararooy. 2010. A Simple and Effective Measure for Complex Low-Level Dependencies. In *Proc. Int'l Conf. Program Comprehension (ICPC)*. IEEE, 80–83.

Rolf Brickenkamp, Lothar Schmidt-Atzert, and Detlev Liepmann. 2010. *Test d2-Revision: Aufmerksamkeits-und Konzentrationstest.* Hogrefe Göttingen.

Julie Brisson, Marc Mainville, Dominique Mailloux, Christelle Beaulieu, Josette Serres, and Sylvain Sirois. 2013. Pupil Diameter Measurement Errors as a Function of Gaze Direction in Corneal Reflection Eyetrackers. *Behavior Research Methods* 45, 4 (2013), 1322–1331.

Ruven Brooks. 1983. Towards a Theory of the Comprehension of Computer Programs. *Int'l Journal of Man-Machine Studies* 18, 6 (1983), 543–554.

Teresa Busjahn, Roman Bednarik, Andrew Begel, Martha Crosby, James H. Paterson, Carsten Schulte, Bonita Sharif, and Sascha Tamm. 2015. Eye Movements in Code Reading: Relaxing the Linear Order. In *Proc. Int'l Conf. Program Comprehension (ICPC)*. IEEE, 255–265.

Siyuan Chen, Julien Epps, Natalie Ruiz, and Fang Chen. 2011. Eye Activity as a Measure of Human Mental Effort in HCI. In *Proc. Int'l Conf. Intelligent User Interfaces*. ACM, 315–318.

Michael J. Doughty. 2001. Consideration of Three Types of Spontaneous Eyeblink Activity in Normal Humans: During Reading and Video Display Terminal Use, in Primary Gaze, and while in Conversation. *Optometry and Vision Science* 78, 10 (2001), 712–725.

Maria K. Eckstein, Belén Guerra-Carrillo, Alison T. Miller Singley, and Silvia A. Bunge. 2017. Beyond Eye Gaze: What Else Can Eyetracking Reveal about Cognition and Cognitive Development? *Developmental Cognitive Neuroscience* 25 (2017), 69–91.

Sarah Fakhoury, Yuzhan Ma, Venera Arnaoudova, and Olusola Adesope. 2018. The Effect of Poor Source Code Lexicon and Readability on Developers' Cognitive Load. In *Proc. Int'l Conf. Program Comprehension (ICPC)*. IEEE, 11 pages.

Benjamin Floyd, Tyler Santander, and Westley Weimer. 2017. Decoding the Representation of Code in the Brain: An fMRI Study of Code Review and Expertise. In *Proc. Int'l Conf. Software Engineering (ICSE)*. IEEE, 175–186.

Denae Ford, Titus Barik, and Chris Parnin. 2015. Studying Sustained Attention and Cognitive States with Eye Tracking in Remote Technical Interviews. *Eye Movements in Programming: Models to Data* (2015), 5 pages.

Thomas Fritz, Andrew Begel, Sebastian C. Müller, Serap Yigit-Elliott, and Manuela Züger. 2014. Using Psycho-Physiological Measures to Assess Task Difficulty in Software Development. In *Proc. Int'l Conf. Software Engineering (ICSE)*. ACM, 402–413.

Michael S. Gazzaniga, Richard B. Ivry, and George R. Mangun. 2013. *Cognitive Neuroscience: The Biology of the Mind.* Norton & Company.

Maurice Halstead. 1977. *Elements of Software Science.* Elsevier Science Inc.

Matthias Hartmann and Martin H. Fischer. 2014. Pupillometry: the Eyes Shed Fresh Light on the Mind. *Current Biology* 24, 7 (2014), R281–R282.

Eckhard H. Hess and James M. Polt. 1964. Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science* 143, 3611 (1964), 1190–1192.

Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures.* OUP Oxford.

Jesper F. Hopstaken, Dimitri van der Linden, Arnold B. Bakker, and Michiel A. J. Kompier. 2015. The Window of My Eyes: Task Disengagement and Mental Fatigue Covary with Pupil Dynamics. *Biological Psychology* 110 (2015), 100–106.

Marcel A. Just and Patricia A. Carpenter. 1993. The Intensity Dimension of Thought: Pupillometric Indices of Sentence Processing. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 47, 2 (1993), 310.

Bruno Laeng, Sylvain Sirois, and Gustaf Gredebäck. 2012. Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science* 7, 1 (2012), 18–27.

Seolhwa Lee, Danial Hooshyar, Hyesung Ji, Kichun Nam, and Heuiseok Lim. 2017. Mining Biometric Data to Predict Programmer Expertise and Task Difficulty. *Cluster Computing* (2017), 1–11.

Keith Nolan, Aidan Mooney, and Susan Bergin. 2015. Examining the Role of Cognitive Load When Learning to Program Program. January (2015), 2–4.

Unaizah Obaidellah, Mohammed Al Haek, and Peter C.-H. Cheng. 2018. A Survey on the Usage of Eye-Tracking in Computer Programming. *ACM Comput. Surv.* 51, 1, Article 5 (Jan. 2018), 58 pages.

Norman Peitek, Janet Siegmund, and André Brechmann. 2017. Enhancing fMRI Studies of Program Comprehension with Eye-Tracking. In *Proc. Int'l Workshop on Eye Movements in Programming*. Freie Universität Berlin, 22–23.

Janet Siegmund, Christian Kästner, Sven Apel, Chris Parnin, Anja Bethmann, Thomas Leich, Gunter Saake, and André Brechmann. 2014. Understanding Understanding Source Code with Functional Magnetic Resonance Imaging. In *Proc. Int'l Conf. Software Engineering (ICSE)*. ACM, 378–389.

Janet Siegmund, Norman Peitek, Chris Parnin, Sven Apel, Johannes Hofmeister, Christian Kästner, Andrew Begel, Anja Bethmann, and André Brechmann. 2017. Measuring Neural Efficiency of Program Comprehension. In *Proc. of the 2017 11th Joint Meeting on Foundations of Software Engineering (ESEC/FSE)*. ACM, 140–150.

John A. Stern, Donna Boyer, and David Schroeder. 1994. Blink Rate: A Possible Measure of Fatigue. *Human factors* 36, 2 (1994), 285–297.