Quality Evaluation of Modern Code Reviews Through Intelligent Biometric Program Comprehension Supplements

Supplement 1.

Psychophysiological Measurements and The Relationship with the Neurophysiology of Cognition

The human nervous system comprises two major subdivisions: the central nervous system (CNS) and the peripheral nervous system (PNS). The PNS is responsible for carrying messages from the CNS to the muscles, glands, organs, and senses, and vice versa. The PNS is subdivided into Somatic Nervous System (SNS) and Autonomic Nervous System (ANS). The SNS is mainly associated with activities of the conscious or voluntary. The ANS, on the other hand, is subdivided into sympathetic and parasympathetic nervous systems. In the sympathetic division, the body tends to activate in response to external stimuli such as stress, distraction. In contrast, the body attempts to relax in the parasympathetic to return to its day-to-day operations [36]. The sympathetic and parasympathetic nervous systems also significantly contribute to brain activity [37].

In our work, we are mainly concerned with the use of Heart Rate Variability (HRV) and task-evoked pupillary response as a surrogate of the cognitive state of software developers while performing code reviewers' tasks. In this context, it is relevant to provide some background on these two essential aspects.

The ANS is responsible for steering blood pressure through the constriction and relaxation of blood vessels and controlling heart rate and contractility. By measuring the heart rate variability (e. g. through the variations of the interval between consecutive R-peaks in the ECG), it is possible to access the ANS sympathetic and parasympathetic dynamics systems. These measures can be obtained in the time and frequency domain, and non-linear methods can be used to access its dynamics. In the time domain, the most common measures used in literature are the standard deviation of NN intervals (SDNN), the root mean square of successive differences between normal heartbeats (RMSSD), the number of pairs of successive NN (R-R) intervals that differ by more than 50 ms (NN50), and the proportion of NN50 divided by the total number of NN (R-R) intervals (pNN50) (see [38] for additional background on this topic). While both the sympathetic and parasympathetic systems contribute to SDNN (which is also highly correlated with frequency bands below 0.14 Hz), the RMSSD is more influenced by the parasympathetic system (and more correlated with high-frequency bands). The pNN50 is also closely related to the parasympathetic system and correlated with the RMSSD and high-frequency band. In the frequency band, the main measurements relate to the quantification of the power and area of the frequency spectrum with specific intervals (very low frequency - VLF: 0.003-0.04 Hz; low frequency - LF:0.04-0.15 Hz and high frequency - HF: 0.15-0.4 Hz) and the ratio of those powers (LF/HF). While the VLF band happens to result from the heart's intrinsic nervous system, the LF power can be associated with the activity of both the sympathetic and parasympathetic systems (and BP regulation via baroreceptors) [38]. The HF band is mainly associated with the parasympathetic system and is also determined by the respiratory rhythms. The ratio between the LH and HF (which we use in our work) has long been understood as a measure of the balance between the sympathetic and parasympathetic systems. Still, it is known that this relationship is not straightforward. Nevertheless, it can still be seen as a measurement of the dominance of one system over the other. [38]

The ANS also has an essential role in modulating pupil size through the innervation of the iris sphincter by the parasympathetic nervous system and the dilator by the sympathetic nervous system, assessing fluctuations in the pupil diameter, which is an important marker of the ANS activity. Contrarily to HRV frequency measurements, pupil diameter variability measurement is not so consensual. While some defined the limit of the low-frequency band on 0.15Hz [39], others determined it to be above this value (e.g., 1.6Hz [40]). Consequently, several values have also been used for the limit of the HF band (0.5Hz [41] and 4Hz [40]). In [41], the authors concluded that the sympathetic system is reflected in the pupil with rhythms around 0.47 Hz. In contrast, the parasympathetic system produced rhythms with slightly higher frequency, around 0,73 Hz (also with higher variability).

Understanding developers' and reviewers' cognitive state and brain mechanisms behind code comprehension and code review require the ability to assess the reviewer's cognitive load. Recent approaches emerged by leveraging neuroscience and heavy instruments such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) [42]. The relevance of EEG and fMRI studies on programmers' mental states results from the fact that they rely on information that comes directly from the CNS and have the potential to be used to validate and calibrate much cheaper, portable, and nearly non-intrusive devices that can also provide (indirect) information on programmers' cognitive states. These biometric techniques use psychophysiological signals driven by the ANS such as HRV, respiratory response, pupillary response (pupillometry), or electrodermal activity (EDA) to assess cognitive load and cognitive states while executing specific tasks, which are well-established mechanisms to evaluate the difficulty and mental effort associated with task execution [35, 43-51].

References

[37]. Beissner, F., Meissner, K., Bär, K. J., & Napadow, V. (2013). The autonomic brain: an activation likelihood estimation meta-analysis for central processing of autonomic function. Journal of neuroscience, 33(25), 10503-10511. doi: 10.1523/JNEUROSCI.1103-13.2013

[38]. F. Shaffer and J.P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms", Front Public Health. 2017 Sep 28;5:258. doi: 10.3389/fpubh.2017.00258. PMID: 29034226; PMCID: PMC5624990.

[39]. A. Murata and H. Iwase, "Evaluation of mental workload by fluctuation analysis of pupil area," in Engineering in Medicine and Biology Society,1998. Proceedings of the 20th Annual International Conference of the IEEE, 1998, pp. 3094-3097.

[40]. V. Peysakhovich, M. Causse, S. Scannella, and F. Dehais, "Frequency analysis of a task-evoked pupillary response: Luminance-independent measure of mental effort," International Journal of Psychophysiology, vol. 97, pp. 30-37, 2015.

[41]. Ana Franco, Carlos Marques Neves, Carla Quintão, Ricardo Vigário, Pedro Vieira, "Singular Spectrum Analysis of Pupillometry Data. Identification of the Sympathetic and Parasympathetic Activity", Procedia Technology, Volume 17, 2014.

[42]. Barbara Weber, Thomas Fischer, René Riedl, "Brain and autonomic nervous system activity measurement in software engineering: A systematic literature review", Journal of Systems and Software, DOI: 10.1016/j.jss.2021.110946, March 2021.

[43]. J. Veltman and A. Gaillard, "Physiological workload reactions to increasing levels of task difficulty," Ergonomics, vol. 41, no. 5, 1998.

[44]. G. F. Walter and S. W. Porges, "Heart rate and respiratory responses as a function of task difficulty: The use of discriminant analysis in the selection of psychologically sensitive physiological responses," Psychophysiology, vol. 13, no. 6, pp. 563–571, 1976.

[45]. B. Pfleging, D. K. Fekety, A. Schmidt, et al., "A model relating pupil diameter to mental workload and lighting conditions," inProceedings of the 2016 CHI conference on human factors in computing systems, 2016. REPEATS 21

[46]. C. Setz, B. Arnrich, J. Schumm, et al., "Discriminating stress from cognitive load using a wearable EDA device," IEEE Transactions on information technology in biomedicine, vol. 14, no. 2, pp. 410–417, 2009.

[47]. K. Kyriakou, B. Resch, G. Sagl, et al., "Detecting moments of stress from measurements of wearable physiological sensors," Sensors, vol. 19, no. 17, p. 3805, 2019.

[48]. Anh Son Le, Tatsuya Suzuki, and Hirofumi Aoki, "Evaluating driver cognitive distraction by eye tracking: From simulator to driving", Transportation Research Interdisciplinary Perspectives journal, Volume 4, March 2020.

[49]. S. C. Muller and T. Fritz, "Using (bio) metrics to predict code quality online," in 2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE). IEEE, 2016, pp. 452–463.

[50]. R. Couceiro, G. Duarte, J. Duraes, et al., "Pupillography as indicator of programmers' mental effort and cognitive overload," in IEEE Int. Conf. on Dependable Systems and Networks – DSN 2019. IEEE, 2019.

[51]. R. Couceiro, R. Barbosa, G. Duraes Duarte, et al., "Spotting problematic code lines using nonintrusive programmers' biofeedback," International Symposium on Software Reliability Engineering, ISSRE, 2019.