Towards Scientific Study of Technical Interviews Using Eye Tracking

Mahnaz Behroozi
Department of Computer Science
North Carolina State University
Raleigh, NC, USA
mbehroo@ncsu.edu

***Abstract*—A technical interview is a stage in industry hiring process for assessing developers' skills. Despite several critics against technical interviews, there is no reported scientific approach to assess their effectiveness. In my work, I first study developers' concerns about technical interviews. Then, I propose a framework to study whiteboard interview in particular and investigate the factors that introduce extraneous challenges for developers. My thesis proposes studying each technical interview style and finding out the factors that negatively affect developers' performance. This may help in propounding interventions and enhancing technical interviews.**

***Keywords*—technical interviews, eye-tracking, cognitive load, data analysis, qualitative analysis**

I. Introduction

Companies worry a lot about hiring the wrong developer but overlook the high cost of missing out good candidates. For example, rejecting a developer with 10+ years of experience and lots of notable contributions in open-source projects who cannot perform well on a whiteboard in front of an interviewer because s/he suffers from impostor syndrome. Also, most developers are not consistent with their performance in all technical interviews they take [1]. As a result, it is important to monitor one's performance at a given interview and apply interventions as needed to help uncover their real skills. This will help companies avoid high cost false negatives.

In my work, I explore developers' concerns about technical interviews through analyses of their failure experiences. Former studies show that in many cases descriptive reports are not solely reliable [2]. Hence, I propose using biometric data as a sanity check to monitor the existence and the effect of developers' concerns on their performance. To check the feasibility of using biometric data I started off with eye-tracking as it is a common means in detecting cognitive load. Several studies in psychology confirm that cognitive load are detectable through different eye-movement measurements such as blink, pupil dilation, fixation, saccade, and others [3]–[5]. Eye-tracking studies are also getting more prevalent in software engineering [6]. After confirming the negative impacts of a criteria on developers' performance in a particular form of technical interview, I will propose interventions to eliminate that criteria. Then I will check if the interventions were effective. My perspective in suggesting this study framework is to check interviewees' cognitive state and to intervene as needed.

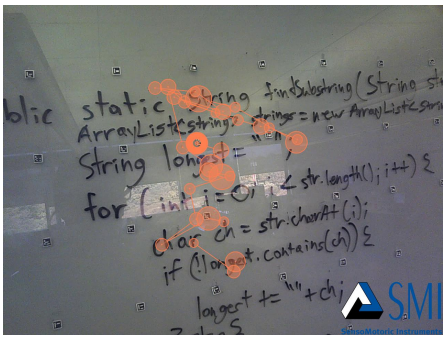

Fig. 1. Fixation scan path of a successful candidate

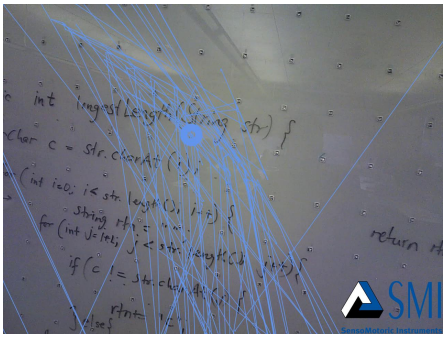

Fig. 2. Fixation scan path of an unsuccessful candidate

II. Preliminary Work and Proposed Study

Phase I. What do developers say about technical interviews? (Completed)

Criticisms about hiring practices are shared as technical interview failure experiences on weblogs, Twitter, social websites such as Hacker News, and etc. To the best of my knowledge, there is no scientific article about these criticisms. In order to support research towards improving technical interviews in software development, I am studying different medias in order to collect developers' experiences on technical interviews. To find developers' main concerns, I conducted a qualitative study through thematic analysis of

small stories on over forty six thousand authored comments from Hacker News with "interview" in their topic [7]. I will extend this part of my study to more inclusive sources, such as Twitter, to get a more detailed picture of developers' concerns about technical interviews.

Phase II. How can we detect differences in cognitive load between different interview settings? (Completed)

We asked our candidates to wear eye-tracker glasses and solve two coding problems of similar difficulty level, one privately on a piece of paper and another one on a whiteboard while a mock interviewer observed them. We hypothesised that the whiteboard setting induces more cognitive load. We found that candidates show more eye-movement characteristics related to elevated cognitive load when they solve the problem on the whiteboard which also match their self-report [8], [9].

Phase III. How does thinking aloud and being observed by an interviewer impact developers' performance? (In progress)

From Phase I of my work, I realized that in whiteboard interviews, *being observed* by an interviewer and *think-aloud* are two bold concerns of the developers. Actually, developers are worried about looking foolish during the interview [10]. This leads me to believe that I can check the effect of these two factors and an intervention to eliminate them by redesigning our experiment in Phase II. Hence, we replaced the paper setting in Phase II with a private whiteboard setting. The aim is to see the effects of making an intervention to eliminate *being observed* and *think-aloud*. Thus, in the private whiteboard setting, in contrast to the public whiteboard setting, our participants were alone in a private room without having to think aloud. In this part of the study we also designed pre-experiment and post-experiment surveys along with NASA Task Load Index (NASA-TLX) [11] questionnaire to document candidates' self-assessment about several aspects of the experiment such as their performance and the cognitive load they experienced during the problem-solving task. We then recruited 70 computer science students consisting of 38 senior undergrads and 32 grads. Our preliminary findings show that successful candidates, passing all the test cases of the problem, have more organized fixation scan paths. Figure 1 and Figure 2 show fixation scan path for a successful and an unsuccessful candidate, respectively.

In this Phase, I am going to specifically work on pupil dilation analysis to identify patterns in both interview settings. I will then proceed with matching patterns with the surveys and candidates' actual performance.

III. Future work

Alternative measurements in addition to eye-movement can offer a deeper insight about the cognitive state of a candidate. I also have our candidates' thinking-aloud voice recordings in the public setting. I will use voice recordings to confirm our findings from eye-movement analysis. In my future studies I also propose integrating heart rate measurement and recording thermal videos during problem-solving in order to have additional means of sanity check.

In addition to the whiteboard interview study, I will keep studying developer concerns on larger set of medias like Twitter, Glassdoor, Reddit, and etc. Once I get a better understanding of the factors causing unnecessary cognitive load in certain forms of interviews, I can design intervention to reduce them. Then, I will study the impact of the interventions.

IV. Conclusion

My dissertation research proposes using eye movement and other biometric measurements to determine the factors that play a significant role in inducing cognitive load to technical interview candidates. The results and findings of this work can suggest better strategies to enhance technical interviews in order to help the candidates show their skills and at the same time reduce false negatives in hiring process.

Acknowledgements. Special thanks to my advisor, Dr. Chris Parnin, for all his insights, support and feedback on this project. I would also want to thank my peers at alt-code research group for their constructive feedback. This material is based in part upon work supported by the National Science Foundation under grant number 1559593.

References

[1] A. Lerner, "Technical interview performance is kind of arbitrary. heres the data." Feb. 2016. [Online]. Available: http://blog.interviewing.io/technical-interview-performance-is-kind-of-arbitrary-heres-the-data/

[2] D. Ford, M. Behroozi, A. Serebrenik, and C. Parnin, "Beyond the code itself: how programmers really look at pull requests," in *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Society*. IEEE Press, 2019, pp. 51–60.

[3] B. Cowley, M. Filetti, K. Lukander, J. Torniainen, A. Henelius, L. Ahonen, O. Barral, I. Kosunen, T. Valtonen, M. Huotilainen *et al.*, "The psychophysiology primer: a guide to methods and a broad review with a focus on human–computer interaction," *Foundations and Trends® in Human–Computer Interaction*, vol. 9, no. 3-4, pp. 151–308, 2016.

[4] S. Chen, J. Epps, N. Ruiz, and F. Chen, "Eye activity as a measure of human mental effort in hci," in *Proceedings of the 16th International Conference on Intelligent User Interfaces*. ACM, 2011, pp. 315–318.

[5] P. Unema and M. Rotting, "Differences in eye movements and mental workload between experienced and inexperienced motor-vehicle drivers." *Visual Search*, pp. 193–202, 1990.

[6] Z. Sharafi, T. Shaffer, B. Sharif, and Y.-G. Guéhéneuc, "Eye-tracking metrics in software engineering," in *2015 Asia-Pacific Software Engineering Conference (APSEC)*. IEEE, 2015, pp. 96–103.

[7] M. Behroozi, C. Parnin, and T. Barik, "Hiring is broken: What do developers say about technical interviews?" in *IEEE Symposium on Visual Languages and Human Centric Computing*. IEEE, 2019, p. to appear.

[8] M. Behroozi, A. Lui, I. Moore, D. Ford, and C. Parnin, "Dazed: measuring the cognitive load of solving technical interview problems at the whiteboard," in *2018 IEEE/ACM 40th International Conference on Software Engineering: New Ideas and Emerging Technologies Results (ICSE-NIER)*. IEEE, 2018, pp. 93–96.

[9] M. Behroozi and C. Parnin, "Can we predict stressful technical interview settings through eye-tracking?" in *Proceedings of the Workshop on Eye Movements in Programming (EMIP'18)*, 2018.

[10] N. Swanner, "Whiteboard interviews absolutely terrify you: Survey," Oct. 2018. [Online]. Available: https://insights.dice.com/2018/10/17/whiteboard-interview-tech-pros-survey/

[11] S. G. Hart, "Nasa task load index (tlx). volume 1.0; paper and pencil package," 1986.