Graduate Research Plan Statement

Background: Functional Neuroimaging – scanning of brain activity – has provided researchers with an understanding of the relationship between activity in certain brain regions and specific mental functions, and is being increasingly applied to study cognitive processes of programming. Not only have these neuroimage studies furthered our understanding of human cognition, but they also have the potential for wider-ranging impacts, such as improving CS pedagogy, technology transfer, and workforce retraining. For example, findings that determined "data structure and spatial operations use the same focal regions of the brain but to different degrees" [1] have led to longitudinal studies reporting positive outcomes on CS test performance of novice programmers when exposed to imaging-derived training [3]. I believe neuroimaging studies have potential positive implications for pedagogy, such as student retention and academic support.

Regardless of the benefits of such findings, medical imaging studies of software are still relatively new and various programming activities have yet to be examined at the cognitive level. Software maintenance and debugging are critical activities (13%-25% of costs [4]), but the programming activity of reading error messages and responding to them, such as by fixing or mitigating the issue, has not been studied at the cognitive level. Given how much of a developer's time is taken by this activity, an extensive amount of research aims to improve the aspect of bug finding, such as localization or error message quality. However, the cognitive processes associated with *interactively reading error messages* and error handling have not been entirely explored. I believe understanding the complex cognitive processes associated with such compounded programming tasks will both provide fundamental knowledge of human processes and also inform the development of new supportive approaches to the overall debugging process.

Proposal: Guided by previous work, **I hypothesize that there is a correlation between spatial and linguistic neural activity and interactive debugging tasks** for two main reasons: (1) debugging tasks often require reasoning about data structures, which has been correlated to spatial neural activity [1]; and (2) a key part of the debugging process consists of reading/writing code and reading of the prose error message, which may be more linguistically focused [7].

I propose a medical imaging study investigating that hypothesis by measuring the neural activity of computer programmers as they respond to compiler error messages. I will recruit groups of undergraduates and graduates from Michigan's computer science department (aiming for at least 38 participants, as per my initial statistical power analysis). All participants will be subjected to medical imaging while completing controlled tasks (using functional near-infrared spectroscopy, or fNIRS). Each stimulus presented to participants will consist of a short Python code snippet, an error location, and an error message. The short Python code snippets will come from successful prior medical imaging work [1,7]. The experimental conditions will consist of (1) varying the programming task (between 3 different types of associated data structures) and, (2) varying the quality of the error message (from unaltered to no semantic information). Based on previous work, I propose a 45-minute interaction with each participant in an IDE framework, admitting responses to 10 programming tasks each.

Evaluation: An important challenge with utilizing functional neural infrared spectroscopy (fNIRS) is the hemodynamic response function. In short, fNIRS monitors hemodynamic response changes of the brain to determine the occurrence of neural activity in a specific brain region. However, these changes may only be detectable for a short period of time and at the initial exposure of the new activity. To best optimize our identification of neural activity and data collection, I propose a behavioral model to structure each task around 4 delineated segments. The segments each represent a subprocess of debugging (reading source code, reading error message, editing source code, and revisiting decisions), with transitions identified via recorded times of all keystrokes and build processes. This will allow for an effective analysis of medical imaging data to check for functional network connectivity or brain activity changes that differ between the groups as the data structure problems and textual error message representations change. Although this proposed evaluation is ambitious, I believe I am well positioned to carry it out effectively: I will build upon the qualitative and quantitative research methods demonstrated in my published peer-reviewed paper and upon the human study techniques described in my under-revision journal article. I will also take advantage of my research group's successful experience with medical imaging-related work.

Planned Extensions: In collaboration with psychology professor Dr. Ioulia Kovelman and members of her research team, we have discussed extending the given proposal by also investigating the neurological activity of dyslexic programmers as they complete debugging tasks. We are interested in investigating how dyslexic and neurotypical programmers debug, especially in the absence of a prose error message. We hypothesize that dyslexic participants may make use of compensating behaviors when reading or writing code because studies have identified semantic compensation as relevant for dyslexic students as they read prose [6]. However, it is unknown if analogous compensation exists for code. We will make use of the psychology lab's established approach for recruiting dyslexic participants.

Intellectual Merit: To the best of my knowledge, I am proposing a novel human study to understand the cognitive processes associated with interactive debugging as it relates to a compiler error message. The closest related work to the proposed study investigates the cognitive processes associated with the decision-making moment of suspicion/bug detection [5]. However, that study mainly focuses on the error detection portion of the debugging process. This prior work did not require participants to interact with an error message or edit code. By contrast, I consider a richer debugging model consisting of four generalized tasks: reading source code, reading prose error message, writing code, and revisiting decisions. My proposed study aims to expand upon prior work by including the processes in which the participant is reading and responding to the error message. By expanding prior knowledge, I will also be contributing new information about how the brain functions in these particular activities.

Broader Impacts: Finding an indicative correlate between spatial or linguistic neural processes and interaction with error messages will help support future efforts in training novice programmers. For example, prior work has indicated success in transferring skills in imaging-identified brain regions to programming skills [3]. I hypothesize that transfer training to improve debugging skills is achievable. Considering debugging is a large aspect for software engineering, it is important to train students in the skills they will need to succeed in their software-related careers. Aside from pedagogical support, our planned extension on studying the effects of error message quality on dyslexic programmers may help provide support and awareness for developers and students with reading impairments. With fields such as software engineering growing in popularity and diversity, it is important to engage in outreach with these communities to better understand their needs and levels of satisfaction, and thus broaden participation.

References:

[1] Y. Huang et al., "Distilling Neural Representations of Data Structure Manipulation using fMRI and fNIRS," International Conference on Software Engineering, 2019.
[2] B. Floyd, T. Santander and W. Weimer, "Decoding the Representation of Code in the Brain: An fMRI Study of Code Review and Expertise," International Conference on Software Engineering, 2017.
[3] M. Endres, M. Fansher, P. Shah, and W. Weimer. 2021. "To read or to rotate: comparing the effects of tenchican reading training and spatial skills training on novice programming ability." Foundations of Software Engineering 2021.
[4] T. Barik et al., "Do Developers Read Compiler Error Messages?," International Conference on Software Engineering (ICSE), 2017.
[5] J. Duraes, H. Madeira, J. Castelhano, C. Duarte and M. C. Branco, "WAP: Understanding the Brain at Software Debugging," International Symposium on Software Reliability Engineering, 2016.
[6] M. Rasamimanana, M. Barbaroux, P. Colé, M. Besson, "Semantic compensation and novel word learning in university students with dyslexia", Neuropsychologia, 2020.
[7] R. Krueger, Y. Huang, X. Liu, T. Santander, W. Weimer and K. Leach, "Neurological Divide: An fMRI Study of Prose and Code Writing,"International Conference on Software Engineering, 2020