An Investigation of the Effects of Mental Fatigue on Programming Tasks' Performance

Saurabh Sarkar
Department of Computer Science
North Carolina State University
Raleigh, NC
ssarkar4@ncsu.edu

Chris Parnin
Department of Computer Science
North Carolina State University
Raleigh, NC
cjparnin@ncsu.edu

ABSTRACT
Mental fatigue reduces the cognitive and physical performance of people. Fatigue has been considered as a crucial issue in different fields like driving but not in software development. We propose to investigate how mental fatigue adversely affects the performance of a software professional, causing systems developed by them to fail in the long run. We have tried answering our research questions and validated our hypothesis with the help of surveys including 311 participants and observing the results of few user studies. This project is to define the research guidelines to further investigate different methods to automatically determine the fatigue level of developers. We can use this information to help developers avoid their fatigue state to control errors in programming, and thereby aid them to be more productive.

Keywords
Software Engineering, Psychology, Fatigue, Work Environment

1. INTRODUCTION
Fatigue is a physiological state of reduced mental or physical mental capability [11]. It results from excessive workload (both physical and mental), exhaustion, or sleeploss. Fatigue is ambiguous in definition. Fatigue (also called exhaustion, tiredness, and lethargy) [15] is a complex phenomenon encompassing several physiological and psychological charactersitics [11]. Fatigue can be categorized according to the distinct effects it has on a person. The inability to perform any physical activities at the level of one's normal activities is *physical/muscular fatigue*. *Mental fatigue* is defined as a state of weariness, with a feeling of boredom/saturation and decline in motivation [9]. Dull emotional responses is characterized as *Emotional Fatigue*. *Skills Fatigue* deals with the inability to perform a certain task at one's usual capacity. People are susceptible to fatigue that could potentially hamper their daily activities by affecting our ability to perform and thus lead to undesirable results.

Mental fatigue is triggered by long and demanding tasks [9]. Exhaustion due to extreme mental and physical work causes performance degradation in execution of human tasks. Mental fatigue also has a negative effect on memory and cognitive functionalities and these functionalities play a vital role in program construction and modeling [3]. Mental fatigue is universal to all software developers [12]. Studies have shown possible relations between mental fatigue and specific programming tasks such as a program construction, modeling and debugging [3].

Primary reasons for system failures are due to low performance of developers, caused by their absentmindedness or exhaustion [5]. Mental fatigue can also lead to system failures in the long run [5]. In today's competitive world, software developers are motivated to work hard, and and hence are involved in numerous projects. Sometimes they work in teams and and at other times, they work individually on a project. All of this leads to sleep deprivation that results in exhaustion and tiredness. Several times developers suffer from psychological diseases which they themselves are not aware of [12]. Developers may make mistakes and therefore introduce bugs during software development. These mental errors are due to reduced cognitive functioning like distraction, low decision making power, less reasoning capabilities, and poor attention and focus [8].

Tests [13] quantify few subjective symptoms of fatigue in performing general tasks, that relates to programming domain, in Table 1.

There is not a lot of work done in this problem area, so our approach would be to create incremental models that can be refined over time with continuous evaluation of the research. Our contribution defines some measurements for measuring the mental fatigue and analyzing its effect on developers' performance, which are discussed further in the paper. This is a step towards working to solve the problem of developers' mental fatigue affecting various programming tasks in the software development process. In Section 2, we illustrate the related works in the field of fatigue and what other cognitive aspects have been researched in programming domain. Section 3 presents how we define some research questions and how they are answered by the help of a survey and a formulative study. Here we also discuss the results of our study and give a detailed analysis of the observations providing a vision of our research. In Section 4, we discuss how the study can be helpful in the field of software engineering.

Table 1: **Subjective symptoms of fatigue**

Exhausted & Drowsy	Mental decline	Incongruity in body & nervous systems
Sleepiness	Nervousness	Physical Strength
Restlessness	Unwillingness	Pain in limbs
Feeling tired	Lack of focus	Strain in eyes
Sluggish	Absentmindedness	Dizziness
Lethargy	Anxiety	
	Weariness	

2. RELATED WORK

Empirical studies have been conducted in other cognitive aspects of software engineering. Khan et. al. [7] have worked on the effect of programmers' moods on the performance of programming debugging, which also comes under the umbrella of psychological causes. Our approach takes mental fatigue as the psychological factor rather than mood and focuses on all programming tasks, not just debugging. Pimenta et. al. [1] [2] have worked on monitoring and analyzing human performance with respect to computer related tasks and detailing the effects of fatigue on performance. Our proposal is the extension of this approach by using different ways to analyze the effects of fatigue and and helping the developer overcome it. Saito [13] has worked on assessment of physical fatigue in industries. Researchers conduct several studies to understand the risk of reduced cognitive capability while driving or performing other physical activities in industries. However, very little research is carried out in assessing the risk for programming. Articles [10] and blogs [12] discuss about the effect of fatigue on the efficacy of developers in tasks like program understanding, construction, modeling, debugging and decision making. We believe that industry requires a tool that can aid a programmer in detecting their state of mental fatigue and help overcome it. This is a first of the many steps taken in this direction, opening a wider scope for more research in the domain.

3. METHODOLOGY

The approach is to test our claim and then set some research guidelines to investigate methods to determine fatigue state of developers. In order to set the research guideline, we have answered two research questions:

- RQ1: *How severely and frequently does fatigue impact developer productivity?*
- RQ2: *What are the factors that lead to mental fatigue and how does it affect a programmer's life?*

We conducted a formative study to answer to these questions. Additionally, the methodology comprises of a study to try setting some benchmark, which includes initial experimental validation using an instrumentation tool *DevFatigue* (Eclipse plug-in).

3.1 Survey

In order to validate the hypothesis of mental fatigue deteriorting performance of developers and to resolve the challenge of quantifying and measuring fatigue with respect to the programming tasks, we intend to build a novel model to classify fatigue and relate it. However, to achieve that, we must set some guidelines on the basis of which we will be able to quantify and define fatigue. This will help us answer the above mentioned research questions.

We conducted a survey and received responses from 311 participants. This survey was distributed with *checkbox.io*[1]. Participants entered in a draw for three $50 gift cards. Links to this survey were posted on reddit groups, Quora, posted in Computer Science Facebook groups, and emailed out to list servs that were directed towards software developers. The age distribution of the participants lies mostly between 17 and 74. Around 50% of the participants lied between the age of 20 to 35, implying that they are aware of the current trend of the industry and of the work culture followed in software companiescompanies, while the others are older, implying experience in software industries. The answers from our respondents helped us gather data about fatigue and investigate possible detection and alleviation mechanisms. The survey consists of three parts: questions about sleep habits and fatigue levels, questions about fatigue and work performance and question about work habits[2].

3.1.1 Results

The results of our survey show developers' view towards fatigue. It helped us identify the factors leading to mental fatigue and fatigue's influence on the performance of any programming task. In addition, several implications allow us to speculate on the way developers behave when they are in the fatigue state.

Six questions were straight forward and objective in nature. We expected direct conclusions from them. The responses helped us answer the research questions:

RQ1: *How severely and frequently does fatigue impact programmer productivity?* One of our survey questions was on a sliding scale asking developers to rate the frequency and severity of the effects of fatigue according to their personal experience. The responses are shown in Figure 1. The responses implied that it is not a major issue but an issue worth looking into. Around 50% of the respondents thought it has a high severity. 45%-50% respondents stated that it occurs to them more often than not. This answers our first research question and motivates us to delve into the problem of mental fatigue of developers in software industry.

RQ2: *What are the factors that lead to mental fatigue state*

[1]http://checkbox.io/
[2]https://github.ncsu.edu/alt-code/fatique/blob /master/study/survey.md

and how does it affect a programmer's life? Based on previous studies [13], both authors came up with some factors that might lead to mental fatigue. We asked the respondents to choose all the factors they think are correct. We ranked the factors according to the responses we received [Figure: 2]. *Stress* and *sleepiness* are the most voted factors, which validates our definition of mental fatigue in programming tasks', that triggers from sleepless nights and day-long works. Physical activities sometimes play a role in triggering exhaustion as well. Khan et. al. [7] have worked on the effect of programmers' moods on the performance of programming debugging, which is one of the fators according to the responses.

To identify the effects of mental fatigue, we focused on a single survey question: *Do you feel when you are tired that it influences your work? If yes, what are some examples?* Both authors randomly selected 30 responses to this question in order to identify initial reasons in examples. We used an online[3] open card sort to group the responses to this question into categories. Card sorting is a technique used to create mental models and define taxonomies from data [4]. Problem label clarifies the theme behind their reasons. We repeated the exercise twice, once individually and the next time together. We compared the labels and finalized Seven labels. The inter-rater agreement score was 21/30 responses. Mostly the disagreement was between *Productivity* & *Performance* and *Mental Work Set* & *Motivation*. The confusion between Productivity and Performance was due to the overlap in the definitions. We define Performance as the speed of completing a task and Productivity as the completion of a task, to help us sort better.

Although all the factors are some what related to each other. But, it can be seen in the responses [Figure: 3], that respondents think that mental fatigue affects Performance the most. This conclusion can be considered as one of the research guidelines. We have conducted some studies to validate this hypothesis and and reached some conclusions in the following sections.

3.1.2 Threats to Validity

Although our survey study provides data on fatigue and its detrimental effects as seen by programmers, there are several threats to validity that should be considered when interpreting our results. First of all, the survey was not controlled and included large age group people from 17 to 74. The survey was posted on various groups assuming that those groups are mostly accessed by software developers but there was no control over the population. There might be people from different roles of software industry who might have different and conflicting opinions. In the responses, there were some respondents as well who answered that fatigue is not an issue. As all the fields of the responses were not mandatory, so some responses were blank. In such a scenario, one cannot be sure that the other fields were answered honestly by such respondents. These type of threasts are mostly common in case of online surveys but a positive fact about the analysis here is that the authors were not biased towards the responses because *checkbox.io* masks the data with anonymous user details before providing the data.

[3]https://conceptcodify.com/

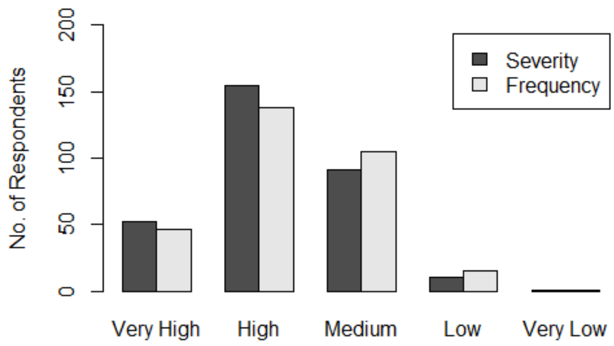

Figure 1: Severity & Frequency of fatigue.

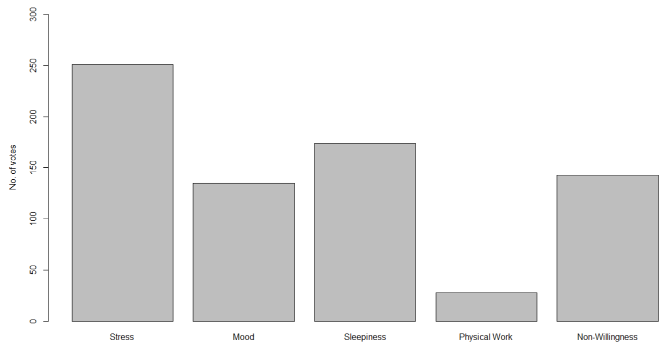

Figure 2: Factors leading to mental fatigue.

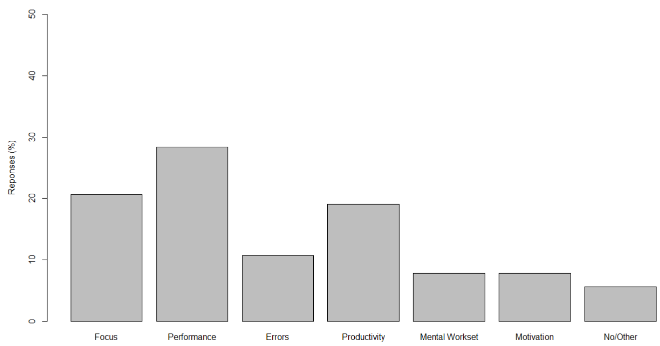

Figure 3: Factors influenced by mental fatigue.

3.2 Observation

To investigate the usefulness of identifying fatigue during programming tasks and recognizing its impact on the performance, we evaluated the outcome of the survey study by conducting some other empirical studies of developers.

We implemented a plug-in for the Eclipse IDE. DevFatigue[4] is an activity tracking plug-in for Eclipse. It is an extension of Rabbit[5]. Like Rabbit, it works in the background with Eclipse and tracks all the activities you perform. It only tracks the actions when Eclipse is active. And logs the data in XML (human readable) format at specific location.

The Rabbit features can be referred to on its official site[5]. Additional features in DevFatigue:

- User Activity - The typing speed (key board usage) of the user with respect to a time period.
- Focus Events - The activities related to keys and mouse usage, like Key Up, Key Down, Mouse Clicks, Mouse Velocity, etc. with specific to time period.
- Project Events - Information regarding the projects like 'imports', and commands used with respect to time period.

The architecture of the proposed framework is the extension of the frameworks used in some of the previous studies [2]. The framework collects biometrics, specifically keystroke and mouse dynamics, that will help in detection and classification of degradation in performance, which can be an effect of fatigue. We are trying to collect the data in a non-intrusive and dynamic way and the indicators of mental fatigue recorded by DevFatigue are in Table

1. Keydown Time: time spent between two consecutive key down and key up events.
2. Errors per Key Pressed: number of times a correction is made per key pressed.
3. Mouse Velocity: velocity of the cursor.
4. Mouse Acceleration: acceleration of the mouse.
5. Time between Keys: time spent between two consecutive key up and key down events.
6. Time between Clicks: time spent between two consecutive mouse up and mouse down events.

All the above mentioned indicators have been proved useful in previous studies [1] [2]. These indicators are collected in a specific time span and provide us information we need to analyze the working patterns of the users and to infer whether the programmer is fatigued or not.

[4]http://www4.ncsu.edu/ ssarkar4/fatigue/eclipse/updatesite/
[5]https://code.google.com/p/rabbit-eclipse/

Table 2: Post-Hackathon Questionnaire

S.No.	Questions
Q1	Age?
Q2	Any industrial experience?
Q3	How would you rate the quality of your code in a scale of 1 to 10, 1 being the least?
Q4	What factors do you think might have affected your performance?
Q5	Hours slept/rested during the hack-a-thon?
Q6	Do you think you could have done better, if you had proper rest in between?
Q7	Did you win?

Table 3: Spearman Correlation Coefficient of features with Energy

Feature	Score
Key Error	0.5247477
Mouse Velocity	0.6903078
Mouse Events	0.7084381
Key Events	0.690758
Navigation	0.6464797

3.2.1 Hack-a-thon

Overnight programming ccompetitions provide a platform for developers to work towards solving a problem. A hackathon is a coding event where computer developers collaborate in building a software product. LexisNexis organized a fall hack-a-thon on 23rd Oct 2014. We asked the participants to take part in our study. All of the participants were graduate students of the Computer Science department at North Carolina State University (hereafter refered as NCSU).

Of the 13 participants invited to participate in the study, only 2 participants finished the study and turned in the data by DevFatigue. We collected data using the log files dumped by DevFatigue, and a post-study questionnaire [Table 3].

We intended to come up with a focus curve [Figure 6] that can show some working pattern of the user and build a model based on the data collected to classify fatigue depending on the user's activities. In our analysis, we defined *Energy* as the number of activities performed in a time span. The graph in Figure 4 shows the working pattern of a user from the Hack-a-thon. The corresponding curves are specified by different colors. The block surrounded by vertical dotted lines represent the sleeping time as shared by the participant. We can observe that all the features are diminishing while approaching the sleep time. We observed that the corelation of Energy with other features were significant [Table 3].

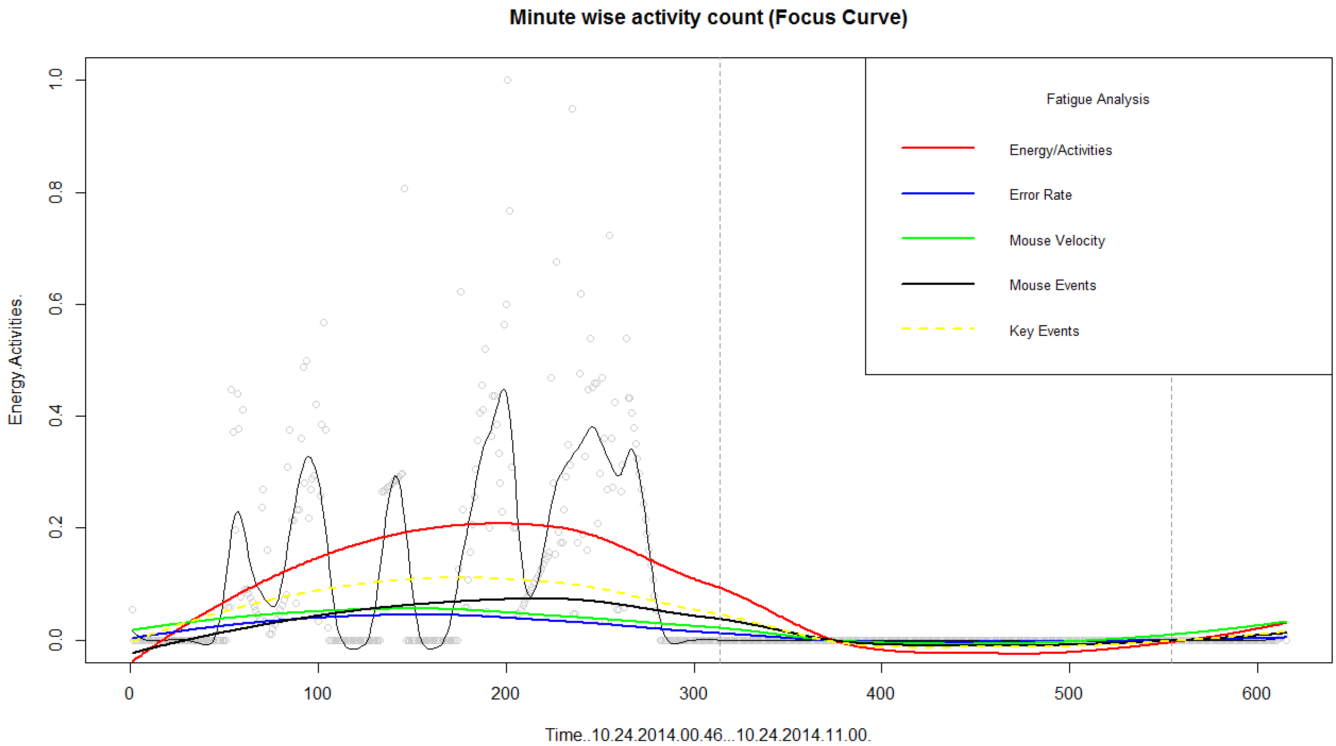

Figure 4: Focus Curve - Hack-a-thon Data.

The results were promising and gave us the motivation to continue our study. Due to lack of more sessions we could not derive a relation between different working patterns of the participant. Hence, we decided to analyse a data set where we could find more working sessions of a user. We used UDC (Usage Data Collector) data set that was available to us.

3.2.2 *Usage Data Collector*

UDC[6] is a framework in Eclipse to collect usage data information. UDC data sets contain the activity information of a particular user, specifically the commands used through a session. More than 50000 records were available for each user. We decided to just analyse 100 days sessions of 2 users. The sessions were short but comparative to other sessions. For each user, we collected the energy score and calculated the average energy score of a particular session (time span). Our baseline approach was to compare day time sessions to night time sessions, with the assumption that one of the session set would have a fatigue effect. We used a threshold of 20 minutes to declare a session different from another. Users show similar working patterns on same days of the week. For comparing sessions, firstly, we considered pairing the scores by day of week. Secondly, we considered session time as the other comparing feature. We paired similar timed sessions. Finally, before processing the data set we normalized all the energy sessions to half-an-hour sessions. We used student's t-test (two tailed) and Cohen's d test for statistically analysing how each session is different from another session. The null hypothesis considered is: H_0: The energy level of day time sessions would be similar to night time sessions of a user. H_1: The session durations of day time sessions would be similar to night time sessions of a user. For each set of samples compared, the test returns a p-value, with a small p-value suggesting that it is unlikely that hypothesis is true. Thus, for the test whose p-value < α the difference is considered to be statistically significant, i.e. hypothesis is rejected. In this work a value of $\alpha = 0.05$ is considered, a standard value generally accepted by research. Table 4 details each hypothesis and its corresponding p-score. The H_0 is rejected as its p-score < α for one user, but not for the other. The rejection of our hypothesis states that there are difference in energy level in day and night sessions. H_1 is accepted for both the users but the score makes us think again about the possibility that it might be rejected for some other sessions. We have compressed the data before processing first by averaging it for a time session and then normalizing, that might have lost some low level information.

Table 4: Statistical analysis of hypothesis depending on different sessions of user

User	Hypothesis	p-score
1	H_0	0.001663433
	H_1	0.19836863
2	H_0	0.959305717
	H_1	0.331089607

The analysis in case of UDC data had some limitations, we

[6]https://eclipse.org/epp/usagedata/

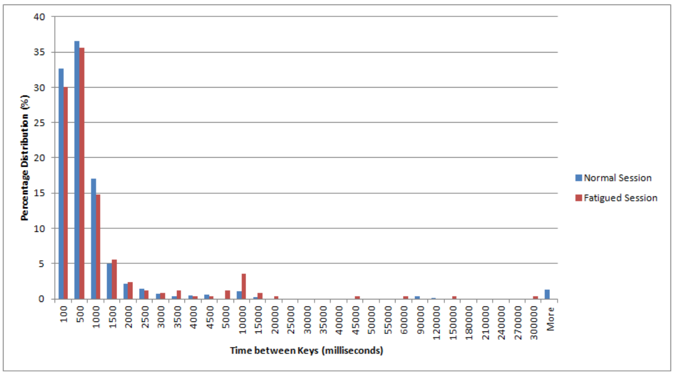

Figure 5: Time between keys distribution difference between a normal and a fatigued session.

had just one feature data, i.e. energy level (activities). Overall the results are interesting to warrant further inspection by using finer data sets and feature sets.

3.2.3 In class study

Dr. Parnin is teaching CSC 510 Software Engineering in Fall 2014 at NCSU. Most students in the course have some experience with coding. Dr. Parnin asked the students to install DevFatigue and let it track their activies on Eclipse, for a couple of weeks. In addition, we asked them to log their daily sleep hours. Of the 100 students enrolled in the course, 19 students responded for the extra credit offered. The data shared by the students had more than one session log and the logs contained detailed time-stamps. We calculated more features and their correlation coefficients and found them significant with respect to energy level. According to sleeping time reported by the students, we compared different sessions of which one can be assumed to be more fatigued than the other.

- *Time between Keys*: Spearman Correlation = 0.7032088
  We observed the time between keys of two sessions and compared their distribution through out the session by taking the frequency of the time in milliseconds. The comparison can be seen in Figure 5. We observed that in a normal session, the short delays covered around 93% but in a fatigued session, the short delays just covered around 85% and the others were long delays. We categorized short delays as 1 to 2000 milliseconds. We can conclude that under the state of mental fatigue, the time between keys increases eventually.

- *Time between Clicks*: Spearman Correlation = 0.737913
  The first thing which we observed were that there were more mouse events (clicks, scrolls, and navigations) in fatigued session than normal session. The other observation was that, in fatigued session the durations were in the same time cluster from 10000 to 30000 milliseconds mostly, whereas in normal session it was scattered. We can conclude that in fatigued state user demonstrates a sluggish mouse usage and it remains constant through out the session and usage of mouse is higher compared to that of keys in fatigued sessions.

- *Mouse Velocity and Acceleration*
  Mouse events in a normal session is less than that of fatigued session as mentioned above. Here we have noticed an opposite behavior as of time between clicks. Velocity and Acceleration in normal session is around the same cluster of speed, but in case of a fatigued session it is scattered. We cannot have a definite conclusion here, but the observation states that in case of fatigue, a user demonstrates more mouse activity and the speed is random but in a normal session his speed is defined.

4. CONCLUSION & FUTURE WORK

The study of mental fatigue, including its causes and symptoms, is traditionally supported by data collected through instrumentation, self-reporting mechanisms (generally questionnaires) or, more recently, through the use of physiological sensors. To comprehend the mental fatigue of software developers we conducted a survey and card sort to catalog factors that lead to fatigue. This study resulted in answering two of the many research questions related to fatigue in software development. The survey helped us figure out the factors which are affected by fatigue and the next step would be to help detect those factors and help developers to overcome the fatigue state. It could be achieved by various alerts, screen freezes or even suggesting Pomodoro[7] technique. The research is to conduct, monitor, and analyze data in a non-invasive and non-intrusive way and present the results in a cordial manner. We hope this paper will encourage similar approaches for finding out how to alleviate mental fatigue. The results of the conducted formative studies motivates us to delve further to find more evidence to support our hypothesis.

Our research aims to help developers avoid their fatigue state to control errors in programming, and eventually aid them for a better productivity. The application would be most suitable for the software industry. We have a lot of different programming tasks in an industry and fatigue might affect the performance/productivity of any of those activities. Thus, we will try conducting some studies on the real environment programming activities. To check the relation between mental fatigue and performance, we plan to review bug and code logs. Our research would be taking the following factors into consideration: working scenarios and behaviors. Smith et. al.[14] studied the effects of intake of coffee on alertness and performance. Coetzer and Richmond [6] performed an empirical analysis on working in teams and its relation to performance.

The survey contained many other questions such as asking developers about their coping strategy under the fatigue state. Future work would include coping mechanism for fatigue. With the increased interest in the behavior of software developers, more research should be carried out to identify the adverse effects of mental fatigue on software development.

5. ACKNOWLEDGEMENT

Thanks to the CSC510 class at North Carolina State University for their participation throughout the process of developing this research. I would like to thank the Alt-Code Group for their support and reviews of the paper.

[7]http://en.wikipedia.org/wiki/Pomodoro_Technique

6. REFERENCES

[1] P. N. André Pimenta, Davide Carneiro and J. Neves. Monitoring mental fatigue through the analysis of keyboard and mouse interaction patterns. *HAIS, LNAI 8073*, pages 222–231, 2013.

[2] P. N. André Pimenta, Davide Carneiro and J. Neves. Analysis of human performance as a measure of mental fatigue. *HAIS, LNAI 8480*, pages 389–401, 2014.

[3] S. B and M. JE. Syntactic/semantic interaction in programmer behaviour: A model and experimental results. *Int J Comput Inf Sci*, 8(3):219–238, 1979.

[4] A. Begel and T. Zimmermann. Analyze this! 145 questions for data scientists in software engineering. In *In Proceedings of the 36th International Conference on Software Engineering (ICSE 2014).*, June 2014.

[5] R. E and I. B.". "the critical role of high-availability infrastructure support services.". Technical report, "Gartner Research, Stamford", "2001".

[6] C. GH and R. L. Investigation of the effects of coffee on alertness and performance during the day and night. *Team Performance Management*, 13(1/2):5–20, 2007.

[7] I. A. Khan and W.-P. Brinkman. Do moods affect programmers' debug performance? *Cogn Tech Work*, 13:245–258, February 2011.

[8] N. M. Larson GE, Alderton DL and U. E. Further evidence on dimensionality and correlates of the cognitive failures questionnaire. *Br J Psychol*, 88(1):29–38, 1997.

[9] V. C. G. N. L. P. J. P. A. P. P. W. Lauren Aaronson, Cynthia Teel and A. Wingate. Defining and measuring fatigue. *Journal Of Nursing Scholarship*, 31(1):45–50, 1999.

[10] H. Makabee. Effective software design, August 2011. http://effectivesoftwaredesign.com/2011/08/23/how-decision-fatigue-affects-the-efficacy-of-programmers/.

[11] P. M. Muchinsky. *Psychology Applied to Work*. Wadsworth/Thomson Learning, Belmont, CA, sixth edition, 2010.

[12] D. Nanghaka. Developer fatigue, August 2012. http://dndannang.blogspot.com/2012/07/developer-fatigue.html.

[13] K. Saito. Measurement of fatigue in industries. *Industrial Health*, 37:134–142, 1999.

[14] F. R. M. A. Smith AP, Brockman P and T. M. An empirical analysis of the relationship between adult attention deficit and effi-cacy for working in teams. *Neuropsychobiology*, 27(4):217–223, 1993.

[15] Wikipedia. Fatigue (medical), November 2014. http://en.wikipedia.org/wiki/Fatigue_(medical).