Sensing Developers' Emotions: The Design of a Replicated Experiment

Daniela Girardi, Filippo Lanubile, Nicole Novielli
University of Bari Aldo Moro
Italy
{daniela.girardi, filippo.lanubile, nicole.novielli}@uniba.it

Davide Fucci
University of Hamburg
Germany
fucci@informatik.uni-hamburg.de

ABSTRACT

Software developers experience a wide variety of emotions during their work and research is now focusing on the role played by these emotions on software developers productivity as well as on their wellbeing. In this paper, we propose a replication of a study aimed investigating to what extent biometric sensors can be used to automatically detect developers' emotions during programming tasks. The long-term goal of our research is to discover which emotions affect developers' productivity and wellbeing during their work. Specifically, we aim at defining approaches for early detection of negative affective states that are known to impair mental wellbeing and productivity. [1]

CCS CONCEPTS

• **Collaborative and social Computing** → Collaborative and social computing theory, concepts and paradigms; *Computer supported cooperative work*; • **Software creation and management → Collaboration in software development**; *Programming teams*

KEYWORDS

Emotion detection, biometric sensors, replicated experiment, empirical software engineering.

ACM Reference format:

D. Girardi, F. Lanubile, N. Novielli, D. Fucci. 2018. Sensing Developers' Emotions: The Design of a Replicated Experiment. In *Proceedings of 3rd International Workshop on Emotion Awareness in Software Engineering* June 2, 2018, Gothenburg, Sweden, 4 pages.
DOI: 10.1145/3194932.3194940

ACM ISBN 978-1-4503-5751-7/18/06…$15.00
https://doi.org/10.1145/3194932.3194940

1 INTRODUCTION

Recent research has shown that developers experience and express a wide range of emotions during collaborative software development [3][16]. Emotions impact work performance when complex cognitive tasks, requiring creativity, are involved [1] such as in software development. Consequently, a research trend emerged to i) study the link between emotions and developers' productivity and software quality [4][5][7][8][14][16][15][22], ii) understand the triggers for emotions at work [4][10], and iii) assess the impact of emotions on the developers' wellbeing [9][13] . In this context, awareness of one's own and others' emotions is the first step towards developing emotional intelligence, the ability to perceive and express emotion, assimilate emotion in thought, understand and reason with emotion, and regulate emotion in the self and others [19].

In a software company, emotion awareness is beneficial for many stakeholders involved in the software development lifecycle. Increasing developers' emotional awareness can be beneficial to improve productivity, resilience to failures and wellbeing. In such scenario, we envisage the development of systems able to detect developers' negative emotions, such as stress or frustration. Our goal is to support them by suggesting corrective actions—e.g., take a break, do mental well-being, exercises using smartphone app like Rize[2]. Furthermore, the team manager or the Scrum master can benefit from the understanding of developers' emotions. For example, they can identify and correct uneven task distribution, support a team member in solving a task, or simply listen to her problems and propose possible solutions. At the organizational level, information about developers' emotional state can be used to evaluate a software development methodology. For a company that promotes and applies Agile development, detecting that most of developers are stressed and frustrated can sign that agile principles are not being applied correctly or that developers are not aligned with such principles, thus increasing the risk of developers' burnout and high turnover. Among the information sources that can be exploited for emotion detection, emotion recognition from biometrics is a consolidated research field [6][11][12][17][20].

[2] http://rizenow.com

In particular, we envisage the use of biometrics collected using low-cost, non-invasive sensors. Non-invasive sensors are crucial in the software development scenario to ensure that they can be worn by software developers without impacting their comfort. In our previous research, we conducted a preliminary study to investigate to what extent low-cost, non-invasive sensors can be used for detecting valence and arousal of emotions triggered by visual stimuli [6].

In this paper, we present the design of the replication of a previous study by Muller and Fritz [14] about investigating the role of emotions in software engineering (Section 2). Consistently with the goal of the original study, we will investigate i) the range of emotions experienced by developers during change tasks; ii) how they correlate with perceived progress; and iii) to what extent they can be automatically detected using biometric sensors. As in the original experimental settings, we will measure signals related to the activity of brain, skin, and heart. We will use the biometric feedback collected through sensors to train and evaluate a machine learning classifier able to distinguish between positive and negative emotions. Moreover, we will investigate the relationship between emotions and developers' perceived progress (Section 3). The findings of this study will complement previous evidence reported by Muller and Fritz about the antecedents of emotions experienced during development tasks. Therefore, this study fits into our long-term goal to uncover the causes behind developers' emotions and the consequences on their mental wellbeing and productivity.

2 THE ORIGINAL STUDY

In the original study that we plan to replicate, Muller and Fritz investigated the relationship between emotions experienced by developers' during a change task and their perceived progress [14]. Using biometric sensors, the authors investigated which aspects and practices affect developers' emotions and suggested good practices to avoid negative emotions and getting stuck. In addition, they distinguished between positive and negative emotions.

Muller and Fritz experiment involved 17 subjects, 11 Ph.D. students with a major in Computer Science and 6 professional software developers'. Participants performed two change tasks while wearing a Neurosky MindBand EEG sensor [3] and an Empatica E3 wristband[4]. The former is a headset for recording electrical brain activity, while latter is a bracelet equipped with a sensor for recording electro-dermal activity (EDA) and a plethysmograph (PPG). While EDA measures the electricity flow in the skin, PPG measures the volume of blood passing through the tissues in a localized area with each heartbeat (or pulse). In addition, they used an Eye Tribe[5] eye tracker to capture measures relate to pupil movement.

[3] http://neurosky.com/
[4] https://www.empatica.com/
[5] http://theeyetribe.com/

Emotiv Insight	Eye Tobii 4c	Empatica E4
(a)	(b)	(c)

Figure 1. Biometric sensors to measure physiological signals: EEG (a), eye tracker (b), EDA and PPG (c)

Each change task lasted for 30 minutes. When a participant showed strong signs of positive or negative emotions, the experimenter interrupted her and asked to rate her emotional state on a scale between -200 (very unpleasant) and 200 (very pleasant), using as reference the two axes of Russell's circumplex model [18]. Contextually, the subjects also rated their perceived progress on a five-point Likert scale. In addition, the experimenter asked to elaborate the reasons for the participant's current emotional state and progress, as well as what could help them to feel better. In the last phase, the experimenter obtained a baseline for the emotion signals by displaying two sets of positive and negative images—before each set, participants watched a two-minutes video of a fish tank. Finally, the experimenter interviewed participants asking in which moments they felt specific emotions and the associated reasons.

Authors performed both quantitative and qualitative analysis of the collected data. They built a supervised machine learning classifier able to identify positive and negative emotions from biometric data with an accuracy of 71,36%. In the qualitative analysis, they found that developers' feel better when they are able to locate a relevant part of the source code and that the action of writing code has a strong impact on developers' perceived progress, independently of its correctness. In addition, they identified good practices to avoid negative emotions and getting stuck: i) switching to a different task; ii) talking to others; iii) taking a break; iv) setting clear goals; and v) allocating sufficient resources ahead of time. Moreover, they reported that negative emotions sometimes can foster developers' productivity. The latter finding is in line with the evidence provided by Wrobel about positive correlation between anger and increase in productivity [22].

3 REPLICATION

As in the original study [14], three main research questions will guide our controlled experiment:

RQ1: What is the range of developers' emotions during change tasks and are developers' emotions correlated with their perceived progress?

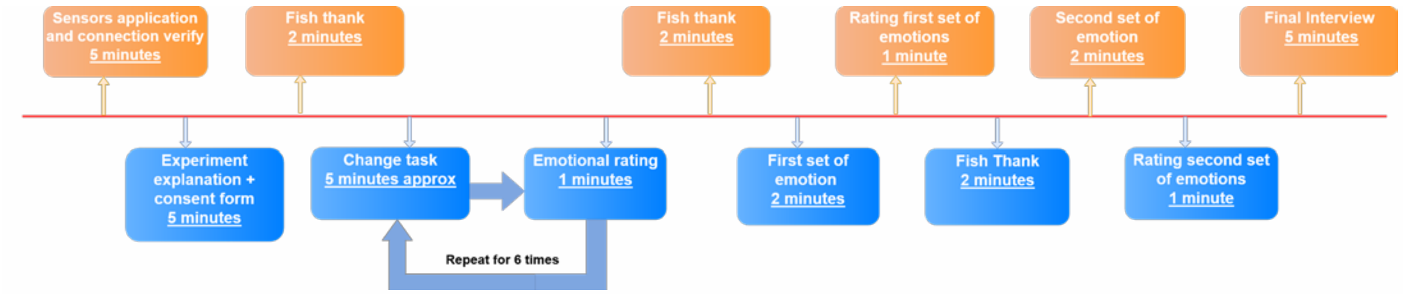

Figure 2. Timeline of the experiment

RQ2: What are aspects and practices that affect developers' emotions and progress during change tasks?

RQ3: Can we use biometric sensors to determine developers' emotion and progress during change tasks?

To assess RQ1, we will replicate one of the change task proposed in the original study. Specifically, we intend to propose the change task consisting in writing a small Java program that interacts with the Stack Exchange API to retrieve all answer posted by a specific user on Stack Overflow and sum up the scores the user earned for these answers. As in the original study, we will investigate RQ2 by mean of questionnaires.

As for biometric sensors, we will use noninvasive, wearable devices comparable to those used in the original study [14]. Specifically, we will adopt the headset Emotiv Insight[6] for EEG measures, Empatica E4 wrist band for EDA and BVP measures and Tobii[7] 4c eye tracking (see Figure 1).

We plan to recruit 30 to 40 subjects for the experiment among students from Software Engineering courses of the Bachelor's degree in computer science at University of Bari (Italy) adopting a sample by saturation strategy. In the original study, Muller and Fritz have compared perceived emotions of Ph.D. students and professional developers founding that the level of expertise does not impact the difference in emotions felt by the subjects. To validate this finding, we plan to further extend our pool of subjects by recruiting participants also among Master's Degree or Ph.D. students.

During the experiment, we will implement the same protocol used in the original study, with the only exception of small adjustments in the emotion self-report approach. Specifically, based on findings from related research, we will collect emotional valence and arousal rates during interruptions in the change task using a smaller range scale, specifically a nine-point scale, through a Self-Assessment Manikin (SAM) mannequin [2] due to its wide application in affective computing [12][20].

[6] https://www.emotiv.com/insight/

[7] https://www.tobii.com/

To address RQ3, we will build a dataset of biometric measurements collected for each subject in presence of a visual stimulus for emotion triggering. In particular, we will capture baseline biometric for emotion reactions to the set of negative and positive images used in the original study. Baseline biometric feedback will be collected also under a neutral emotional condition for all subjects—i.e., by showing them a two-minute fish tank video.

The total duration of the experiment will cover no more than 60 minutes. In Figure 2, we report the detailed timeline regarding the phases of the experiment.

4 CONCLUSION

The main goal of this proposed replication is to collect evidence to support or refute the findings, reported in the original study, about the relationship between emotions felt during a software development task and developers' perceived progress. Additionally, since a need to investigate the role of negative emotions emerged in the original study, we will explore under which circumstances they can be beneficial or detrimental to developers' progress.

ACKNOWLEDGMENTS

This work is partially funded by the project: 'EmoQuest - Investigating the Role of Emotions in Online Question & Answer Sites', funded by MIUR (Ministero dell'Università e della Ricerca) under the program "Scientific Independence of young Researchers" (SIR), the project 'Riconoscimento di emozioni basato su sensori biometrici' funded by the Department of Computer Science of the University of Bari (ex 60% 2015), and by the project OPENREQ, funded by the Horizon2020 program of the European Union. Project Id: 732463

REFERENCES

[1] T.M. Amabile, Sigal G. Barsade, Jennifer S. Mueller, Barry M. Staw. 2005. Affect and Creativity at Work. *Administrative Science Quarterly*, Volume 50, 367–403.

[2] M. M. Bradley and P. J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. Journal of Behavioral Therapy and Experimental Psychiatry, Volume 25, 49–59.

[3] F. Calefato, F. Lanubile, N. Novielli. 2018. How to ask for technical help? Evidence-based guidelines for writing questions on Stack Overflow. *Information and Software Technology*, Volume 94, February 2018, 186–207

[4] D. Ford and C. Parnin. 2015. Exploring causes of frustration for software developers. In *Proceedings of the Eighth International Workshop on Cooperative and Human Aspects of Software Engineering* (CHASE '15). IEEE Press, Piscataway, NJ, USA, 115–116.

[5] T. Fritz, A. Begel, S. C. Müller, S. Yigit-Elliott, and M. Züger. 2014. Using psycho-physiological measures to assess task difficulty in software development. In *Proceedings of the 36th International Conference on Software Engineering* (ICSE 2014). ACM, New York, NY, USA, 402–413.

[6] D. Girardi, N. Novielli, F. Lanubile. 2017. Emotion Detection Using Noninvasive Low Cost Sensors. In Proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII 2017), 125–130.

[7] D. Graziotin, X. Wang, and P. Abrahamsson.2013. Are happy developers more productive? the correlation of affective states of software developers and their-self-assessed productivity. In Proceedings of the 14th International Conference on Product-Focused Software Process Improvement (PROFES 2013), 50–64.

[8] D. Graziotin, X. Wang, and P. Abrahamsson. 2014. Happy software developers solve problems better: psychological measurements in empirical software engineering. PeerJ, vol. 2, e289

[9] D. Graziotin, Fabian F., Xiaofeng Wang, and P. Abrahamsson. 2017. Consequences of unhappiness while developing software. In *Proceedings of the 2nd International Workshop on Emotion Awareness in Software Engineering* (SEmotion '17). IEEE Press, Piscataway, NJ, USA, 42–47.

[10] D. Graziotin, F. Fagerholm, X. Wang, and P. Abrahamsson. 2017. On the Unhappiness of Software Developers. In Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering (EASE'17). ACM, New York, NY, USA, 324–333.

[11] Z. Guendil, ,Z. Lachiri , Z. Maaoui Z., and A, Pruski. 2015. Emotion recognition from physiological signals using fusion of wavelet based features. In Proceedings of the 7th International Conference on Modelling, Identification and Control (ICMIC), 1–6.

[12] S. Koelstra., C. Muhl, M. Soleymani, J. S. Lee, A.Yazdani, T. Ebrahimi, T. Pun, A.. Nijholt, and I. Patras. 2012. Deap: A database for emotion analysis; using physiological signals. IEEE Transactions on Affective Computing, Volume 3, 18-31.

[13] M. Mäntylä, B. Adams, G. Destefanis, D. Graziotin, and M. Ortu. 2016. Mining valence, arousal, and dominance: possibilities for detecting burnout and productivity? In *Proceedings of the 13th International Conference on Mining Software Repositories* (MSR '16). ACM, New York, NY, USA, 247–258.

[14] S. C. Müller and T. Fritz. 2015. Stuck and frustrated or in flow and happy: sensing developers' emotions and progress. In *Proceedings of the 37th International Conference on Software Engineering* - Volume 1 (ICSE '15), Vol. 1. IEEE Press, Piscataway, NJ, USA, 688–699.

[15] S. C. Müller and T. Fritz. 2016. Using (bio)metrics to predict code quality online. In Proceedings of the 38th International Conference on Software Engineering (ICSE '16). ACM, New York, NY, USA, 452–463.

[16] A. Murgia, P. Tourani, B. Adams, and M. Ortu. 2014. Do developers feel emotions? an exploratory analysis of emotions in software artifacts. In Proceedings of the 11th Working Conference on Mining Software Repositories (MSR 2014). ACM, New York, NY, USA, 262–271.

[17] M. Murugappan and S. Murugappan. 2013. Human emotion recognition through short time electroencephalogram (eeg) signals using fast fourier transform. In Proceeding of the IEEE 9th International Colloquium on Signal Processing and its Applications (CSPA 2013), 289294.

[18] J. A. Russell. 1980. A Circumplex Model of Affect. Journal of Personality and Social Psychology, Volume 33, 1161–1178

[19] P. Salovey and J. D. Mayer. 1990. Emotional Intelligence. Imagination, Cognition and Personality, Volume 9, 185–211.

[20] M. Soleymani, M. Pantic and T. Pun. 2012. Multimodal emotion recognition in response to videos. IEEE Transactions on Affective Computing, Volume 3, pp. 211–223.

[21] X. Wang, D. Nie, and B. Lu. 2014. Emotional state classification from EEG data using machine learning approach. Neurocomput. 129 (April 2014), pp.94-106.

[22] M. Wrobel, Emotions in the software development process.2013. In Proceeding of the 6th International Conference on Human System Interaction (HSI 2013), 518–523.