DRAFT 2020-09-23
The latest version is available at erik.bjareholt.com/thesis/goaldocument.pdf

M.Sc. Thesis proposal

Classifying brain activity using low-cost biosensors and automated time tracking

Erik Bjäreholt
(erik@bjareho.lt, dat13ebj@student.lu.se)

Student	Erik Bjäreholt
Supervisor	Markus Borg
Examiner	Elizabeth Bjarnarson
Start date	1st October
End date	1st March

CONTENTS

NOTES & TODOS

I. BACKGROUND

People spend more time than ever using computing devices. Work, entertainment, and services, have been steadily moving online over the last few decades and this trend is expected to continue. While several studies have been tracking how people spend time on their devices a wider study of how people's app usage is changing over time and how it varies with demographics, is not publicly available.

Furthermore, how different device activities affect the user behaviorally and neurologically is of interest for many areas of research, including:

- psychological well-being, such as depression and social anxiety [1][2], stress [3], self-esteem, life satisfaction, loneliness, and depression [4].
- the impact of screen time on children and adolescents [5].
- attention span among media multitasking adults [3].
- enhancing personal productivity [6].

Understanding device use and the underlying cognitive processes are essential when designing for motivation, engagement and wellbeing in digital experiences [7].

This becomes especially relevant for knowledge workers, such as software developers, who spend the majority of their working time on computing devices.

A. Automated time trackers

Automated time-trackers have been developed for computing devices for various applications such as tracking productivity, managing excessive use of social networking sites (SNSs).

1) Commercial use: Companies like RescueTime [8], Hubstaff [9], and others offer automated time tracking as a service. These services let the user track their screen time by installing a program on their device which tracks the active application and sends the data to their servers for storage and analysis. The user can then view their data in a dashboard on the service's website. Some of these services, like RescueTime and Hubstaff, are marketed towards teams and professionals, who want to keep track of individual and team productivity.

However, these services have some issues for use by researchers and individuals alike. Notably, their collection of detailed and non-anonymized behavioral data into a centralized system bring significant privacy concerns, especially in cases where the data is shared with a team or an employer.

Other limitations of these services, such as low temporal resolution and limited event detail, cause additional issues for certain tasks that are timing-sensitive (such as ERPs), or preprocessing steps that can take advantage of high level of detail (like classifying activity).

2) Research use: Previous research has been published which used automated time trackers, such as TimeAware [6] and ScreenLife [10]. However, these previous contributions are — like the commercial services — not open source nor permissively licensed, and therefore not available for external research use nor further development.

3) ActivityWatch: The free and open source automated time tracker ActivityWatch [11] addresses aforementioned issues with other software around source availability/licensing, privacy, temporal resolution, event detail, and cross-platform support.

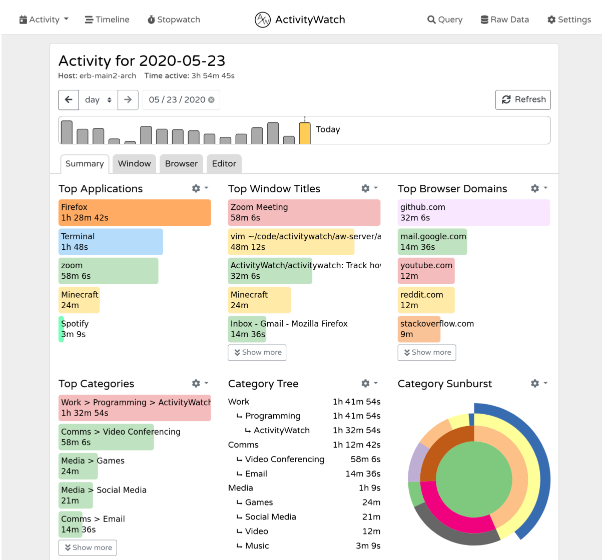

Fig. 1. ActivityWatch activity dashboard. Showing top applications, window titles, browser domains, and categories.

B. Low-cost functional brain imaging

Functional brain imaging methods such as fMRI, fNIRS, and EEG, have been used to study the relationship between cognitive or physical activity, and brain activity [12][13][14]. The more accurate methods such as fMRI are costly and inflexible/impractical for many uses.

However, the recent availability of low-cost biosensors such as EEG, HEG, and fNIRS, enables studying brain activity during real-life tasks. As an example it has been shown that it is possible to classify what task a participant is undertaking using fMRI [12], which has been replicated using EEG and low-cost biosensors [14].

But they are not without their limitations — among them a notably low signal-to-noise ratio [15] — yet visual evoked potentials (VEPs) have been shown to be sufficient for high-speed BCI applications [16].

To combat the low signal-to-noise ratio, machine learning methods have been employed with varying degrees of success. Examples from previous research include Convolutional Neural Networks (CNNs), which have been successful in classifying time series in general [17], and EEG data in particular [18]. As well as Hierarchical Convolutional Neural Networks (HCNNs), which have been used for EEG-based emotion recognition [19].

II. Problem Description, Research Goals and Questions

Knowledge workers in general, and software developers in particular, have varying degrees of productivity that is in part mediated by varying degrees of focus (or "flow"). Physical devices which use device activity as an indicator of flow have been developed to reduce interruptions at work, and subsequently improve productivity [20].

Since EEG and other low-cost biosensors have been used to classify developers emotions [21] and comprehension tasks [14] we will investigate if they can also be used to classify which device activity the user is engaging in.

The goal of this project is therefore to investigate whether EEG and other low-cost biosensors can be used to improve the productivity of developers by studying the relationship between device activity and brain activity. This will be done in part by training a device activity classifier from EEG data.

We structure our goals according to the Goal-Question-Metric (GQM) paradigm. The goals are grouped into phases, which are explained in the Methodology section.

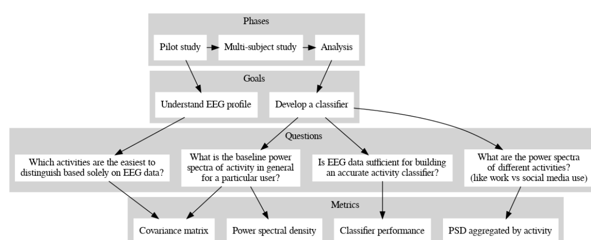

Fig. 2. Overview of the goals/phases (with a subset of the questions and metrics), structured according to the Goal-Question-Metric (GQM) paradigm.

Goal: Characterize the EEG profile of different device activities, to inform the design of the multiple-subject experiment.

- *Question:* Which device activities are the easiest to distinguish from each other based solely on EEG data?
  - *Metrics:* PSD, covariance matrix.
- *Question:* Roughly how much data is required to characterize the EEG of a particular activity?
  - *Metrics:* Time for aggregated power spectra to converge.
- *Question:* Which electrode placements (in the 10-20 system) are most suitable for the task?

  - *Metrics:* Covariance matrix.
- *Question:* What is the baseline power spectra of device activity in general for a particular user?
  - *Metrics:* Power spectra.
- *Question:* What are the power spectra of different activities (like work vs social media use)?
  - *Metrics:* Power spectra aggregated by device activity,
- *Question:* Does the same activity consistently yield similar power spectral densities.
  - *Metrics:* Variance of the aggregated power spectra, both per user and across the entire group.
- *Question:* Are any brain regions more strongly associated with certain activity? (similar to G2-Q3)
  - *Metrics:* Source estimation, covariance matrix.

Goal: Develop a classifier for device activity from EEG data.

- *Question:* Is the EEG data sufficient for building a accurate classifier for device activity?
  - *Metrics:* Classifier performance
- *Question:* What classifier is suitable for the dataset?
  - *Metrics:* Accuracy, precision, F1 score of different classifiers.
- *Question:* Which electrode placements/brain regions have the most predictive power?
  - *Metrics:* Source estimation, covariance matrix.

A. Research Questions

Can low-cost biosensors, like EEG, be used to...

- RQ1. Can input from low-cost EEG sensors be used to train a classifier separating software developers' device activities?
- RQ2. How does the number of channels impact the classification accuracy?

B. Challenges

- Low volume of EEG data collected (limited time for data collection)
- Limitations of low-cost EEG equipment (small number of channels)
- Orthogonal stimuli (eye movement/blinking, use of keyboard/mouse) which will contribute significant noise to the EEG readings not relevant for the classification task.

III. Methodology

We will simultaneously collect EEG and device activity data from subjects during device use. EEG data will be collected with the 8-channel OpenBCI Cyton biosensing board and Ultracortex headset. Device activity data will be collected and categorized with ActivityWatch. The activity categories will be used to label the dataset so we can train the classifier.

Project planning and tasks will be managed using GitHub issues and the Kanban-like board provided by GitHub (GitHub Projects).

The project can be roughly split into three phases.

- Run a single-subject pilot study and perform preliminary analysis.
- Design and run a multiple-subject study.
- Develop a classifier for device activity.

A. Pilot study

The purpose of the first phase is to collect a pilot dataset (using a single subject) to inform the design of the main study.

The main questions we want to answer in this phase are about the study design in phase 2. Such as to identify which activities are the easiest to distinguish using EEG and are thus best suited for further study. (See GQM for more questions to be answered)

B. Multiple-subject study

The purpose of this phase is to collect data from controlled experiments with multiple subjects.

The experimental design and protocol will be partially drawn from previous research, including a study on recognizing developers emotions while programming [21], along with a replication study utilizing EEG to classify code comprehension [14].

C. Develop classifier

Using the data collected from the previous phase we will develop a classifier that classifies the category of device activity using EEG data.

Additionally, there are numerous available EEG datasets and challenges on Kaggle which will be used to inform the development of the classifier.

IV. Scientific contributions

- Developing a classifier for device activity using low-cost biosensors.
- Identifying relationships between device activity and brain activity, as measured by EEG.
- Validating previous research on biofeedback in software engineering.
- Developing a framework for using ActivityWatch in research.

V. Resources

I have access to all the resources I need to perform the project, including:

- OpenBCI Ultracortex headset and Cyton biosensing board (8 channels)
- Muse S EEG headset (4 channels)
- Access to significant CPU & GPU compute capacity (helpful for machine learning)
- 3–5 male test subjects who have given informed consent.
- ActivityWatch, an open source automated time tracker (already developed by the author, but never before used in a scientific publication)
- Resources from the NeuroTechX community, such as code notebooks.[22]
- Open source software for EEG research such as MNE-Python [23] and Brainflow [24].
- Publicly available EEG datasets on Kaggle.[25]
- Access to researchers who work with EEG at Lund University (at Department of Automatic Control and Department of Psychology)
- Computer

References

[1] Maarten H. W. Selfhout et al. "Different types of Internet use, depression, and social anxiety: The role of perceived friendship quality". In: *Journal of Adolescence* 32.4 (Aug. 1, 2009), pp. 819–833. ISSN: 0140-1971. DOI: 10.1016/j.adolescence.2008.10.011. URL: http://www.sciencedirect.com/science/article/pii/S0140197108001218 (visited on 05/13/2020).

[2] Dhavan Shah et al. "Nonrecursive Models of Internet Use and Community Engagement: Questioning Whether Time Spent Online Erodes Social Capital". In: *Journalism & Mass Communication Quarterly* 79.4 (Dec. 1, 2002). Publisher: SAGE Publications Inc, pp. 964–987. ISSN: 1077-6990. DOI: 10.1177/107769900207900412. URL: https://doi.org/10.1177/107769900207900412 (visited on 05/13/2020).

[3] Gloria Mark, Yiran Wang, and Melissa Niiya. "Stress and multitasking in everyday college life: an empirical study of online activity". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. Toronto, Ontario, Canada: Association for Computing Machinery, Apr. 26, 2014, pp. 41–50. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557361. URL: https://doi.org/10.1145/2556288.2557361 (visited on 05/22/2020).

[4] Chiungjung Huang. "Time Spent on Social Network Sites and Psychological Well-Being: A Meta-Analysis". In: *Cyberpsychology, Behavior, and Social Networking* 20.6 (June 1, 2017). Publisher: Mary Ann Liebert, Inc., publishers, pp. 346–354. ISSN: 2152-2715. DOI: 10.1089/cyber.2016.0758. URL: https://www.liebertpub.com/doi/10.1089/cyber.2016.0758 (visited on 05/13/2020).

[5] Kaveri Subrahmanyam et al. "The impact of computer use on children's and adolescents' development". In: *Journal of Applied Developmental Psychology*. Children in the Digital Age 22.1 (Jan. 1, 2001), pp. 7–30. ISSN: 0193-3973. DOI: 10.1016/S0193-3973(00)00063-0. URL: http://www.sciencedirect.com/science/article/pii/S0193397300000630 (visited on 05/13/2020).

[6] Young-Ho Kim et al. "TimeAware: Leveraging Framing Effects to Enhance Personal Productivity". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: Association for Computing Machinery, May 7, 2016, pp. 272–283. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858428. URL: https://doi.org/10.1145/2858036.2858428 (visited on 05/13/2020).

[7] Dorian Peters, Rafael A. Calvo, and Richard M. Ryan. "Designing for Motivation, Engagement and Wellbeing in Digital Experience". In: *Frontiers in Psychology* 9 (2018). Publisher: Frontiers. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2018.00797. URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00797/full?report=reader (visited on 05/23/2020).

[8] *RescueTime: Automatic Time-Tracking Software*. Library Catalog: www.rescuetime.com. URL: https://www.rescuetime.com/ (visited on 05/23/2020).

[9] *Hubstaff: Time Tracking and Productivity Monitoring Tool*. Library Catalog: hubstaff.com. URL: https://hubstaff.com/ (visited on 05/23/2020).

[10] John Rooksby et al. "Personal Tracking of Screen Time on Digital Devices". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: Association for Computing Machinery, May 7, 2016, pp. 284–296. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858055. URL: https://doi.org/10.1145/2858036.2858055 (visited on 05/13/2020).

[11] *ActivityWatch: Open source automated time tracker*. In collab. with Erik Bjäreholt and Johan Bjäreholt. original-date: 2016-04-27T15:26:09Z. May 23, 2020. URL: https://github.com/ActivityWatch/activitywatch (visited on 05/23/2020).

[12] Benjamin Floyd, Tyler Santander, and Westley Weimer. "Decoding the Representation of Code in the Brain: An fMRI Study of Code Review and Expertise". In: *2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE)*. 2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE). ISSN: 1558-1225. May 2017, pp. 175–186. DOI: 10.1109/ICSE.2017.24.

[13] Keum-Shik Hong, Noman Naseer, and Yun-Hee Kim. "Classification of prefrontal and motor cortex signals for three-class fNIRS–BCI". In: *Neuroscience Letters* 587 (Feb. 5, 2015), pp. 87–92. ISSN: 0304-3940. DOI: 10.1016/j.neulet.2014.12.029. URL: http://www.sciencedirect.com/science/article/pii/S0304394014009744 (visited on 05/22/2020).

[14] Davide Fucci et al. "A Replication Study on Code Comprehension and Expertise using Lightweight Biometric Sensors". In: *2019 IEEE/ACM 27th International Conference on Program Comprehension (ICPC)*. 2019 IEEE/ACM 27th International Conference on Program Comprehension (ICPC). ISSN: 2643-7171. May 2019, pp. 311–322. DOI: 10.1109/ICPC.2019.00050.

[15] D. J. McFarland and J. R. Wolpaw. "EEG-based brain–computer interfaces". In: *Current Opinion in Biomedical Engineering*. Synthetic Biology and Biomedical Engineering / Neural Engineering 4 (Dec. 1, 2017), pp. 194–200. ISSN: 2468-4511. DOI: 10.1016/j.cobme.2017.11.004. URL: http://www.sciencedirect.com/science/article/pii/S246845111730082X (visited on 05/13/2020).

[16] Martin Spüler. "A high-speed brain-computer interface (BCI) using dry EEG electrodes". In: *PLoS ONE* 12.2 (Feb. 22, 2017). ISSN: 1932-6203. DOI: 10.1371/journal.pone.0172400. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5321409/ (visited on 05/13/2020).

[17] Bendong Zhao et al. "Convolutional neural networks for time series classification". In: *Journal of Systems Engineering and Electronics* 28.1 (Feb. 2017). Conference Name: Journal of Systems Engineering and Electronics, pp. 162–169. ISSN: 1004-4132. DOI: 10.21629/JSEE.2017.01.18.

[18] Robin Tibor Schirrmeister et al. "Deep learning with convolutional neural networks for EEG decoding and visualization". In: *Human Brain Mapping* 38.11 (2017). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.23730, pp. 5391–5420. ISSN: 1097-0193. DOI: 10.1002/hbm.23730. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.23730 (visited on 05/22/2020).

[19] Jinpeng Li, Zhaoxiang Zhang, and Huiguang He. "Hierarchical Convolutional Neural Networks for EEG-Based Emotion Recognition". In: *Cognitive Computation* 10.2 (Apr. 1, 2018), pp. 368–380. ISSN: 1866-9964. DOI: 10.1007/s12559-017-9533-x. URL: https://doi.org/10.1007/s12559-017-9533-x (visited on 05/22/2020).

[20] Manuela Züger et al. "Reducing Interruptions at Work: A Large-Scale Field Study of FlowLight". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: Association for Computing Machinery, May 2, 2017, pp. 61–72. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025662. URL: https://doi.org/10.1145/3025453.3025662 (visited on 06/13/2020).

[21] Daniela Girardi et al. "Recognizing Developers' Emotions while Programming". In: *arXiv:2001.09177 [cs]* (Jan. 24, 2020). DOI: 10.1145/3377811.3380374. arXiv: 2001.09177. URL: http://arxiv.org/abs/2001.09177 (visited on 09/23/2020).

[22] *NeuroTechX/eeg-notebooks*. original-date: 2020-08-11T00:25:01Z. Sept. 20, 2020. URL: https://github.com/NeuroTechX/eeg-notebooks (visited on 09/21/2020).

[23] *mne-tools/mne-python*. original-date: 2011-01-28T03:31:13Z. Sept. 12, 2020. URL: https://github.com/mne-tools/mne-python (visited on 09/13/2020).

[24] *brainflow-dev/brainflow*. original-date: 2018-07-15T12:42:26Z. Sept. 13, 2020. URL: https://github.com/brainflow-dev/brainflow (visited on 09/13/2020).

[25] *Search: EEG | Kaggle*. URL: https://www.kaggle.com/search?q=eeg (visited on 09/03/2020).

Further Reading

John Rooksby et al. "Personal tracking as lived informatics". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. Toronto, Ontario, Canada: Association for Computing Machinery, Apr. 26, 2014, pp. 1163–1172. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557039. URL: https://doi.org/10.1145/2556288.2557039 (visited on 05/13/2020).

Tsung-Sheng Chang and Wei-Hung Hsiao. "Time Spent on Social Networking Sites: Understanding User Behavior and Social Capital". In: *Systems Research and Behavioral Science* 31.1 (2014). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sres.2169, pp. 102–114. ISSN: 1099-1743. DOI: 10.1002/sres.2169. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sres.2169 (visited on 05/13/2020).

Yu Zhang et al. "Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces". In: *Expert Systems with Applications* 96 (Apr. 15, 2018), pp. 302–310. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2017.12.015. URL: http://www.sciencedirect.com/science/article/pii/S0957417417308291 (visited on 05/13/2020).

Ji Peng Koh. "Brain computer interface via EEG signals". In: (2018). Accepted: 2018-04-23T02:21:24Z. URL: https://dr.ntu.edu.sg//handle/10356/73955 (visited on 05/13/2020).

Lamyaa Sadouk. "CNN Approaches for Time Series Classification". In: *Time Series Analysis - Data, Methods, and Applications* (Nov. 5, 2018). Publisher: IntechOpen. DOI: 10.5772/intechopen.81170. URL: https://www.intechopen.com/books/time-series-analysis-data-methods-and-applications/cnn-approaches-for-time-series-classification (visited on 05/22/2020).

Christian Mühl et al. "A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges". In: *Brain-Computer Interfaces* 1.2 (Apr. 3, 2014). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/2326263X.2014.912881, pp. 66–84. ISSN: 2326-263X. DOI: 10.1080/2326263X.2014.912881. URL: https://doi.org/10.1080/2326263X.2014.912881 (visited on 05/22/2020).

Ji Young Jung et al. "Caffeine Maintains Arousal Level and Prevents Change of Electroencephalogram Spectral Powers with Time at Rest". In: *Journal of Korean Sleep Research Society* 11.1 (June 30, 2014). Publisher: Korean Sleep Research Society, pp. 5–10. ISSN: 1738-608X, 2288-4912. DOI: 10.13078/jksrs.14002. URL: http://www.e-jsm.org/journal/view.php?number=171 (visited on 08/03/2020).

PeterPutman. *The Effects of a Single Administration of a Moderate Dose of Caffeine on Cognitive Control and Spontaneous EEG Theta/Beta Ratio*. Clinical trial registration NCT02940808. submitted: October 17, 2016. clinicaltrials.gov, Mar. 7, 2017. URL: https://clinicaltrials.gov/ct2/show/NCT02940808 (visited on 07/31/2020).

Craik A, He Y, and Contreras-Vidal Jl. *Deep learning for electroencephalogram (EEG) classification tasks: a review*. Journal of neural engineering. ISSN: 1741-2552 Issue: 3 Publisher: J Neural Eng Volume: 16. June 2019. DOI: 10.1088/1741-2552/ab0ab5. URL: https://pubmed.ncbi.nlm.nih.gov/30808014/ (visited on 09/21/2020).

Roy Y et al. *Deep learning-based electroencephalography analysis: a systematic review*. Journal of neural engineering. ISSN: 1741-2552 Issue: 5 Publisher: J Neural Eng Volume: 16. Aug. 14, 2019. DOI: 10.1088/1741-2552/ab260c. URL: https://pubmed.ncbi.nlm.nih.gov/31151119/ (visited on 09/21/2020).

Andrew Melnik et al. "Systems, Subjects, Sessions: To What Extent Do These Factors Influence EEG Data?" In: *Frontiers in Human Neuroscience* 11 (2017). Publisher: Frontiers. ISSN: 1662-5161. DOI: 10.3389/fnhum.2017.00150. URL: https://www.frontiersin.org/articles/10.3389/fnhum.2017.00150/full (visited on 09/21/2020).

Timeflux. URL: https://timeflux.io/ (visited on 09/21/2020).