Mental Workload as Personal Data: Designing a Cognitive Activity Tracker

Max L. Wilson
Natalia Sharon
Horia A. Maior
Mixed Reality Lab,
School of Computer Science,
University of Nottingham, UK
max.wilson@nottingham.ac.uk
horia.maior@nottingham.ac.uk

Serena Midha
Horizon CDT,
University of Nottingham, UK
serena.midha@nottingham.ac.uk

Michael P. Craven
NIHR MindTech MedTech Co-operative,
Institute of Mental Health &
Bioengineering Research Group,
Faculty of Engineering,
University of Nottingham, UK
michael.craven@nottingham.ac.uk

Sarah Sharples
Human Factors Research Group,
Faculty of Engineering,
University of Nottingham, UK
sarah.sharples@nottingham.ac.uk

Submitted to the CHI2018 Workshop on Computing and Mental Health.

Abstract
Research continues to correlate physical signals with mental activity, as opposed to physical activity, with physiological sensors. Further, with the proliferation of wearable technology, it seems imminent that our smart watches can soon keep track of our mental activity as well as our physical activity. Our research is working towards accurately measuring Mental Workload 'in the wild' using physiological sensors. While we work towards that goal, however, we have begun to explore the design aspects of representing personal cognitive data to users; analogous to a step counter for physical activity. We present the results of diary studies, focus groups, and prototyping exercises to identify design considerations for future cognitive activity trackers.

Author Keywords
Mental Workload; personal data; activity monitoring.

Introduction
Many people are increasingly being encouraged to try and stay *cognitively* healthy, whether to avoid mental decline associated with ageing or to combat degenerative conditions [5,15], or simply to increase effectiveness in daily living, through cognitively stimulating activities [6], physical exercise [13] and by generally managing both work and rest e.g. improving sleep quality [1]. In other cognitive and emotional

Various commercial products are becoming available. https://www.myfeel.co/ estimates emotional changes from a wrist band and https://spire.io estimates stress from breathing rates.

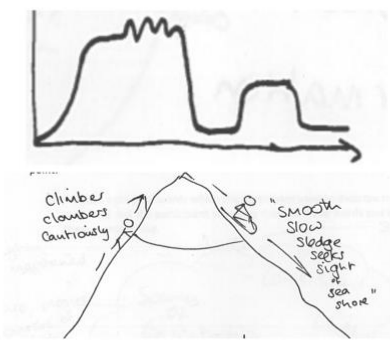

Figure 1: Continuous Representations

Figure 2: Discrete Representations

areas, apps exist to help people monitor their own well-being through frequent self-assessments e.g. of mood [4] or to keep a health diary e.g. for headaches [8]. Furthermore, neurobiofeedback techniques such as EEG, have been studied as a potential therapy in a variety of clinical areas including cognitive function [2].

Traditionally, self-assessment scales have been the most reliable and tested industry methods for measuring Mental Workload (MWL; the amount of mental effort required to complete a task) [7,16] and Emotional Response [3]. While EEG has been used to directly observe MWL in the brain [13], our own work has aimed to use Functional Near-Infrared Spectroscopy (fNIRS), which is tolerant in contexts with higher ecological validity [11]. Maior et al [12] recently used this technique to give people concurrent mid-task feedback on their MWL. More recently, however, much research seeks to identify cognitive and emotional changes by correlating physiological signals, such as MWL with heart rate variability [10], and stress through wrist-worn galvanic skin response [17]. These developments, along with emerging products, highlight that people will soon be able to monitor their own cognitive activity in the same way as physical activity.

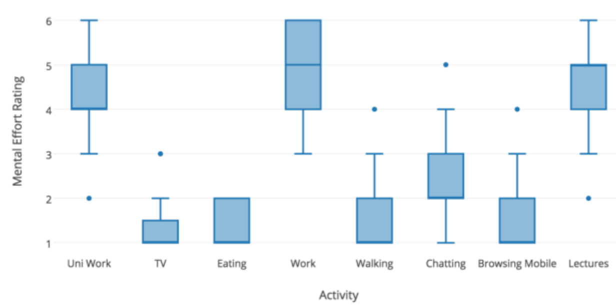

Figure 3: A sample of activities with high and low MWL ratings

This paper reports on research focused on designing such a 'fitbit for the brain'. While guidelines exist for personal informatics [9], cognitive data is very different from physical activity data. We ask: 1) What metaphors do people use when imagining MWL? 2) How do people use those metaphors to evaluate their day? And 3) How should MWL be best visualised for users to gain valuable insights into their cognitive activity?

Stage 1: Diary Studies of Mental Activity

A diary study was designed to capture natural retrospective participant conceptualisations of their MWL. 12 participants were asked to fill out a structured diary template every hour (during waking hours) for three days, which involved a) describing their main activities during the hour, b) identifying what caused the highest and lowest levels of MWL, c) rating those levels out of 6, and d) describing the experience of MWL for those activities. Participants were then invited to a concluding interview to review their diary entries.

Diary Study Results

Participants returned a total of 482 diary entries (~40 each); some example ratings are included in Figure 3. During the interviews, participants described high MWL as: '*When I have a heap of information in my head at once'* and *'trying to juggle lots of thoughts, sometimes feel shaky/nervous'.* When discussing activities rated 5-6 in the diaries, participants used descriptors such as *'concentration', 'stressed'* and *'cluttered'.* Activities which resulted in the lowest mental workload levels of 1-2 were browsing social networks, watching TV and playing game consoles. The majority of these tasks were performed after 4pm and although being rated low, would be the activity resulting in the highest mental workload for that hour. Words used to describe a low mental workload state when performing activities

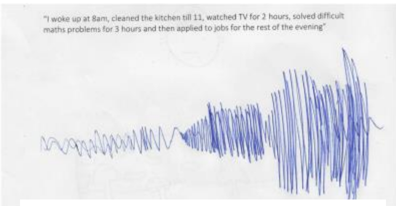

Figure 4: Focus Group: Scenario drawing exercise.

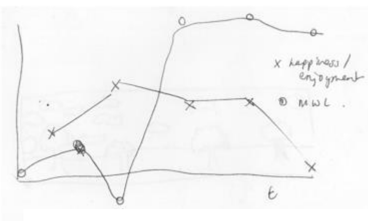

Figure 5: Enjoyment vs MWL.

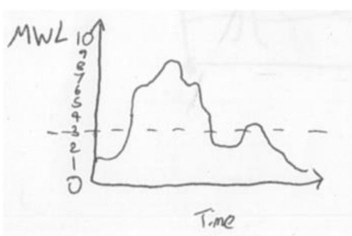

Figure 6: Having a baseline shouldn't be too low or too high.

such as watching TV, or day-dreaming when walking home, included *'auto pilot'* and *'zoned-out'.* Getting organised, planning for the day, and replying to emails tended to occur between 7- 9am, and was rated 2-3 by younger participants, but 4-5 for older participants.

Some participants described MWL with more continuous metaphors (Figure 1), explaining it as "*something on a spectrum which changes throughout the day*". One said "*high mountain peaks of mental effort*" were common where "*the steepness of walk reflects rapid change intensity".* 30% of participants associated high MWL with the colour red, and low with blue; one participant directly used a thermometer metaphor as a scale. A second group of metaphors were more typically discrete concepts (Figure 2), that referred e.g. to being able to handle a fixed number of tasks. Some suggested they had a capacity for a number of things they could *"juggle"*, whilst one participant described their brain as filled *"with too many task bubbles".*

Stage 2: Focus Groups

Four focus groups, including some participants from stage 1, were organised to discuss design ideas in more detail. Focus groups went through four initial stages to develop a deeper consideration for design concepts: 1) discussing what MWL is, 2) listing activities for high and low MWL, 3) interpreting diagrams drawn from stage 1, and 4) drawing diagrams to represent example scenarios (Figure 4). Participants then discussed ideas at a deeper level to identify key design considerations.

Positive and Negative Mental Workload
Groups discussed both merits and concerns with having high MWL. A member of group 1 (see Figure 5) said *"I had two lines – one for mental workload and one for general enjoyment of the situation. I think they're two different things and having a high mental workload doesn't mean you're not enjoying the task – I like maths problems but it can be pretty difficult and so I put them on two separate lines"*. This discussion highlighted an important design consideration: that productive high MWL was considered a positive aspect when working on a task, but prolonged high MWL was typically considered a bad thing. Likewise, prolonged low MWL was considered a negative, but most wished to have periods of low MWL between high activity.

Baselines and Targets for Mental Workload
In relation to positive and negative MWL, both baselines and targets were used to highlight what was good and bad. Some participants considered that the ideal mental activity was not 'low', but around a low-to-middle amount of activity (Figure 6), where users should enjoy periods of rest below the baseline and work productively above it. Participants could then set targets relevant to their day, or indeed for times of day like during the morning, and in the evening.

Counting both high and low MWL
Rather than focusing on how high the MWL was, example targets were given that involve having a good distribution of MWL throughout the day. Compared to number to steps, a potential measure would be number of minutes in high, medium, and low MWL states.

Stage 3: Prototypes

Our more recent work has focused on developing usable prototypes for objectively and subjectively recording MWL data from participants. Figure 7 shows a functional prototype that uses galvanic skin response and heart rate data from a Microsoft band to estimate MWL. Initial machine learning models have been generated to estimate MWL state, which can also be

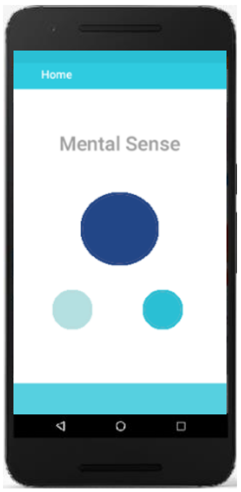

Figure 8: A 'current view' shows share of low, medium and high MWL.

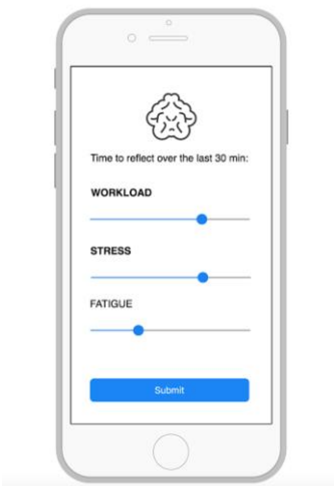

Figure 7: A diary app to support long term subjective data collection.

calibrated for the individual. This algorithm, and indeed the data from the Microsoft band, currently have limited accuracy. The prototype, however, embodies our design recommendations for a cognitive activity tracker, and can be demonstrated at the workshop.

1) The live view shows a distribution of high, medium, and low MWL for the last 20 minutes, using size.
2) Prolonged status in a single state causes the colour to fade (indicating that it is 'worn out').
3) A diary view combines colour codes on a timeline with entries from the participant's calendar, as well as any personal notes.
4) A history view shows overall distributions of low, medium, and high MWL for each day.
5) A configuration section allows users to set their ideal low, medium, and high MWL distribution, and calibrate the app with an N-back test.

Ongoing and Future Work

An ongoing challenge for this type of research, is that there is no clear 'ground truth' that can be used for a user's current MWL level. Subjective techniques are either retrospectively summative, or intermittent and add MWL to the participant. Consequently, machine learning algorithms don't have a clear target to aim for. One thread of our work continues, however, to take increasingly longitudinal in-the-wild measures of oxygenation changes in the pre-frontal cortex using fNIRS; our new fNIRS sensor is completely wireless. Physiological measures from wearable wristbands, taken at the same time, can then be correlated with a form of 'objective ground truth' produced by the fNIRS sensor. Further, we are embarking on a series of studies to take longer-term subjective readings of MWL in daily life using mobile diary apps (Figure 8).

Conclusions

In anticipation that estimating cognitive activity with wearable technology is an imminent possibility, as we do now with physical activity, we have embarked on a series of design research exercises. Using a diary study, interviews, and focus groups, we have collected examples of reflections on the Mental Workload associated with activities in everyday life. We have begun to identify key design considerations for cognitive tracking apps. In contrast to counting amount of physical activity, we expect that participants would benefit more from understanding how their time is distributed across different high, medium, and low Mental Workload states. This would allow users to benefit from knowing when they have worked hard and when they have taken a break from it. Further, setting targets for how these states are distributed would also allow e.g. people worried about mental decline to aim for higher levels of activity throughout the day. Our ongoing work is focused on a) using prototypes to evoke more detailed insights into everyday Mental Workload, b) taking objective measures of Mental Workload in the wild during longer periods, and c) evaluating machine learning approaches to estimate this objective data from wearable technology.

Acknowledgements

Prototypes were developed by Tjaart Broodryk. We acknowledge the financial support of the NIHR MindTech MedTech Co-operative, and the EPSRC (EP/L015463/1, EP/M000877/1).

References

1. Saeed Abdullah, Mark Matthews, Elizabeth L. Murnane, Geri Gay and Tanzeem Choudhury. 2014. Towards circadian computing: "early to bed and

early to rise" makes some of us unhealthy and sleep deprived. In *Proc. UbiComp'14*. 673-684.
2. Jean Alvarez, Fremonta L. Meyer, David L. Granoff and Allan Lundy. 2013. The Effect of EEG Biofeedback on Reducing Postcancer Cognitive Impairment. *Integrative Cancer Therapies,* 12, 6, 475-487.
3. Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25, 1, 49-59.
4. E. Bethan Davies, Michael P. Craven, Jennifer L. Martin and Lucy Simons. 2017. Proportionate methods for evaluating a simple digital mental health tool. *Evidence-Based Mental Health*, 20, 4.
5. J. Antonio Garcia-Casal, Andrea Loizeau, Emese Csipke, Manuel Franco-Martin, M. Victoria Perea-Bartolome and Martin Orrell. 2016. Computer-based cognitive interventions for people living with dementia: a systematic literature review and meta-analysis. *Aging & Mental Health*. 21, 5, 454-467.
6. Global Council on Brain Health. 2017. *Engage Your Brain: GCBH Recommendations on Cognitively Stimulating Activities*.
7. Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*. 52, 139–183.
8. Amos S Hundert, Anna Huguet, Patrick J McGrath, Jennifer N Stinson and Mike Wheaton. 2014. Commercially Available Mobile Phone Headache Diary Apps: A Systematic Review. *JMIR mHealth and uHealth,* 2, 3 (Jul-Sep 2014), e36.
9. Ian Li, Anind Dey and Jodi Forlizzi, 2010. A stage-based model of personal informatics systems. In *Proc. CHI'10.* 557-566.
10. Julia C. Lo, Emdzad Sehic and Sebastiaan A. Meijer. 2017. Measuring Mental Workload with Low-Cost and Wearable Sensors: Insights into the Accuracy, Obtrusiveness, and Research Usability of Three Instruments. *Journal of Cognitive Engineering and Decision Making*, 11, 4, 323-336.
11. Horia A. Maior, Matthew Pike, Sarah Sharples and Max L. Wilson. 2015. Examining the Reliability of Using fNIRS in Realistic HCI Settings for Spatial and Verbal Tasks. In *Proc. CHI'15*. 3039-3042.
12. Horia A. Maior, Max L. Wilson, and Sarah Sharples. 2018. Workload Alerts - Using Physiological Measures of Mental Workload to Provide Feedback during Tasks. *ACM Trans. Comput.-Hum. Interact.* (in press), 30 pages.
13. Joseph Michael Northey, Nicolas Cherbuin, Kate Louise Pumpa, Disa Jane Smee and Ben Rattray. 2018. Exercise interventions for cognitive function in adults older than 50: a systematic review with meta-analysis. *British Journal of Sports Medicine,* 52, 154-160.
14. Winnie K. Y. So, Savio W. H. Wong, Joseph N. Mak and Rosa H. M. Chan. 2017. An evaluation of mental workload with frontal EEG. *PLoS ONE,* 12, 4, e0174949.
15. Yaakov Stern. 2012. Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurology*, 11, 11 (November 2012), 1006–1012.
16. Andrew J. Tattersall and Penelope S. Foord. 1996. An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics,* 39, 5, 740-748.
17. María Viqueira Villarejo, Begoña García Zapirain and Amaia Méndez Zorrilla. 2012. A stress sensor based on Galvanic Skin Response (GSR) controlled by ZigBee. *Sensors,* 12, 5, 6075-6101.