Sleep Quality Tracking and Alignment with a Simple In-the-Wild Task

Andrew Vargo
Osaka Metropolitan University
Sakai, Japan
awv@omu.ac.jp

Masayuki Fujimoto
Osaka Prefecture University
Sakai, Japan
fujimoto@m.cs.osakafu-u.ac.jp

Koichi Kise
Osaka Metropolitan University
Sakai, Japan
kise@omu.ac.jp

ABSTRACT

The introduction of wearable sleep trackers, such as the Oura Ring, allow users to monitor the duration and quality of their sleep. For researchers in the fields of education and cognitive sciences, these wearables promise a new source of valuable data, especially if the wearables can be used outside of the laboratory. A problem in validating the wearable data for daily life is that it is difficult to give tasks which are both effective enough challenging users, but simple and quick enough that they can be completed by participants daily. In this poster, we present and in-the-wild study which employs a simple arithmetic task. The results indicate that sleep quality data from a wearable sleep trackers can be used to gauge performance.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**.

KEYWORDS

Sleep Sensing, In-the-Wild, Readiness

ACM Reference Format:
Andrew Vargo, Masayuki Fujimoto, and Koichi Kise. 2022. Sleep Quality Tracking and Alignment with a Simple In-the-Wild Task. In *Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp/ISWC '22 Adjunct), September 11–15, 2022, Cambridge, United Kingdom.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3544793.3560359

Figure 1: The 100 Square Addition Application.

1 INTRODUCTION

The development of wearable sleep trackers allows users to gain insight into both how long and restorative their sleep actually was. The Oura Ring is an attractive device since it is relatively unobtrusive and offers multi-stage sleep tracking capabilities [1]. One possible goal of using such sleep quality data is to measure the impact sleep has on individuals and build applications that help people manage workloads, rest, and study sessions. If the data can be reliably recorded in-the-wild, then it should be possible to build real-world applications.

One problem that researchers face is in setting tasks that cause enough cognitive load, while still being a task that participants are willing to do repeatedly. Therefore, it is vital to identify tasks that give insight into cognitive statuses while not being intrusive on participant time and effort. For instance, previous work used a 1-minute psychomotor vigilance tasks in an in-the-wild setting to infer the circadian rhythm [2]. In this paper, we implement a simple arithmetic game on a mobile application. The results show that performance can be predicted by certain sleep quality metrics.

2 TASK OVERVIEW

The task we selected was Japanese arithmetic game called "100 Square Calculation" (100SC) [1]. The goal of this game is to always add the coordinates of a 10x10 grid together as fast as possible with as few mistakes as possible. As shown in Figure 1, we used an mobile application that tracked the results and time of completion. The right side of the figure shows a game that was completed in 329 seconds. The six red boxes indicate mistakes, giving a correctness score of 94.

3 EXPERIMENT OVERVIEW

16 Japanese University Students (M=15, F=1, Avg. Age=21.6, StD of Age = 2.18) participated in the study. Participants were volunteers from a pool of university students (N=60) who received Generation 2 Oura Rings. At the time of the study, participants had been wearing the ring for 2-5 months. Participants received a remuneration of 3,000 JPY (Approximately 25 USD). Ethical approval was granted

[1] https://www.shogakukan.co.jp/books/09104425

UbiComp/ISWC '22 Adjunct, September 11–15, 2022, Cambridge, United Kingdom

ACM ISBN 978-1-4503-9423-9/22/09.
https://doi.org/10.1145/3544793.3560359

Table 1: Participant 100 Square Addition Results (Correct out 100 and Duration in Seconds)

ParticipantID	Results	Avg. Correct	StD. of Correct	Avg. Duration	StD. Duration
1	17	98.47	1.74	264.00	24.01
2	12	98.67	1.15	400.33	65.52
3	13	98.62	1.26	286.54	47.71
4	14	98.71	2.40	469.86	57.31
5	12	96.42	2.19	551.42	55.77
6	14	97.50	1.61	341.14	33.57
7	12	96.75	2.14	470.00	35.70
8	14	99.00	1.24	348.07	34.64
9	12	97.67	2.10	504.75	90.53
10	10	93.92	2.72	309.62	18.32
11	14	97.29	1.86	517.43	69.29
12	12	96.00	2.83	409.83	64.67
13	14	94.64	1.74	326.71	46.44
14	15	97.73	1.91	308.53	35.84
15	14	98.29	1.64	409.07	23.49
16	12	98.50	1.73	275.58	26.92

before the experiment began. Participants were asked to both wear the Oura Ring continuously and to take the the 100 Square calculation Test every morning upon awakening. The participants were asked to continue the study for a minimum of 14 days. As Table 1 shows, the maximum number of results submitted was 17, while the minimum was 10.

The 100SC task contains two important elements: time to completion and correctness. Based on Table 1, we estimated the fastest possible completion time was about 200 seconds (In the data set, Min=224 seconds and Max=745 seconds). Correctness (Min=88, Max=100, Mode=99) showed that the task was quite easy for most participants. A correlation test showed that correctness and time to completion were uncorrelated. Since both elements are vital to the performance of the task, we created the Adjusted Score is calculated by combining the amount of time that taken to complete the task and the number of mistakes. The score is fit on a scale were the best possible score is 0 and higher scores worse. $(1 - (Duration/200) + (100/Correcness)/10)$. Thus, the Adjusted Score heavily penalizes mistakes.

4 RESULTS AND CONCLUSION

We fit a linear mixed model with Adjusted Score as the dependent variable and the sleep duration measurements and temperature deviation as the fixed variables. Models were reduced with step-wise reduction, and the best performing model, as shown in Table 2, was selected with AIC. The results show that increased REM Sleep Duration decreases the Adjusted Scores, while increased Temperature Deviation increases the Adjusted Scores. These results are intuitive since REM Sleep is associated with mental restoration and increased Temperature is associated with lower quality sleep. The marginal and conditional pseudo-R2 values were 0.02 and 0.72 respectively, indicating that the sleep scores had a small, but significant, impact on the Adjusted Score.

The results are promising as a simple task was able to give insight into the sleep quality in an in-the-wild setting. This is important for harnessing wearables for building applications that can aide lifestyle changes. Future studies will focus on expanding the pool of participants to include greater age and gender diversity and more focus on the impact of different sleep stages on cognitive activities in-the-wild.

Table 2: Linear Mixed Model Results (*p<0.05)

Variable	Estimate	StD. Error	p-value
Deep Sleep Duration	2.645e-05	1.461e-05	0.072
REM Sleep Duration	-3.413e-05	1.497e-05	0.024*
Temperature Deviation	3.166e-01	1.225e-01	0.011*

ACKNOWLEDGMENTS

This work was supported in part by grants from JST CREST (JPMJCR20G3, JPMJCR16E1), JSPS Grant-in-Aid for Scientific Research (20H04213, 20KK0235), and Grand challenge of the Initiative for Life Design Innovation (iLDi).

REFERENCES

[1] Marco Altini and Hannu Kinnunen. 2021. The Promise of Sleep: A Multi-Sensor Approach for Accurate Sleep Stage Detection Using the Oura Ring. *Sensors* 21, 13 (Jan. 2021), 4302. https://doi.org/10.3390/s21134302 Number: 13 Publisher: Multidisciplinary Digital Publishing Institute.

[2] Benjamin Tag, Tilman Dingler, Andrew W. Vargo, and Vassilis Kostakos. 2020. Inferring Circadian Rhythms of Cognitive Performance in Everyday Life. *IEEE Pervasive Computing* 19, 3 (July 2020). https://doi.org/10.1109/MPRV.2020.2994914