Not Just a Matter of Accuracy: A fNIRS Pilot Study into Discrepancy Between Sleep Data and Subjective Sleep Experience in Quantified-Self Sleep Tracking

Zilu Liang[1,2](✉)

[1] Ubiquitous Computing and Personal Informatics Lab, Kyoto University of Advanced Science (KUAS), Kyoto, Japan
liang.zilu@kuas.ac.jp
[2] Institute of Industrial Science (IIS), The University of Tokyo, Tokyo, Japan

Abstract. Quantified-self sleep tracking devices such as Fitbit are gaining great popular in recent years. However, users often complain about the discrepancy between the data collected with sleep trackers and their subjective sleep experience, which is often attributed to the accuracy issue of the devices. In this pilot study, we aim to provide an explanation to such discrepancy from a neuroscience perspective. We investigated the associations of subjective sleep rating and Fitbit measured sleep data to cortical hemodynamics in the prefrontal cortex (PFC) during the first sleep cycle. Correlation analysis results showed that subjective sleep rating mainly correlated to the median of the concentration changes in oxyhemoglobin (ΔO2Hb) and deoxyhemoglobin (ΔHHb) in a set of channels, with positive correlation coefficients. In contrast, the sleep score computed by Fitbit mainly correlated to the mean of the ΔO2Hb and ΔHHb in a different set of channels, with negative correlation coefficients. The findings suggested that better perceived sleep quality may be positively associated to increased hemodynamics during the first sleep cycle, and the opposite may be true for objective sleep metrics such as sleep score measured by Fitbit. The result implies that users' subjective perception of sleep and the sleep tracking devices may be capturing different dimensions of sleep. As such, improving device accuracy may help little in addressing the discrepancy between the subjective sleep experience and the objective data. The findings provided design implications for the development of future sleep tracking technologies.

Keywords: Quantified-Self · Sleep tracking · Fitbit · fNIRS

1 Introduction

Sleep health has far-reaching effects on people's mental and physical health [1]. The past decade has witnessed significantly advances in quantified-self sleep tracking technologies that allow users to monitor their sleep at home and over an extended time

S. Spinsante et al. (Eds.): HealthyIoT 2021, LNICST 432, pp. 74–87, 2022.
https://doi.org/10.1007/978-3-030-99197-5_7

span. Popular devices like Fitbit and Apple Watch allow users to measure clinically significant sleep variables with reasonable accuracy [2, 3], and have been increasingly used in research studies to measure sleep outcomes for better ecological validity [4–6]. These devices use embedded accelerometer and photoplethysmogram (PPG) to infer the sleep stages that a user goes through at night. Many studies have investigated how users interact with sleep tracking technologies in various forms (e.g., wristbands, headbands, smartphone apps) [4, 7–9], and found that users often complained about the discrepancy between the data collected with these devices and their subjective sleep experience [7, 8]. Such discrepancy also often causes cognitive dissonance that reduces users' trust of the devices and drive them to discontinue their use [7].

The discrepancy between the data measured with sleep trackers and users' subjective sleep experience is often referred to as an accuracy issue, which leads researchers to assume that improving device accuracy is the solution to the problem. Nonetheless, sleep science studies found that the discrepancy between the sleep data and the subjective sleep experience may persist even when the gold standard sleep measurement technique—polysomnography (PSG)—is used, suggesting that such discrepancy is likely to be less a matter of device accuracy but rather a common phenomenon in human sleep. Prior sleep research studies show that sleep metrics derived from PSG data explained poorly the variance in subjective sleep quality [10, 11]. A recent study on how users interact with sleep tracking technology argues that there is no one-on-one mapping between the objective sleep metrics and the subjective sleep quality, despite that users tend to establish a false connection between their subjective sleep experience and some sleep metrics [7]. For example, sleep metrics such as micro-arousals, deep sleep, and REM sleep can barely be consciously perceived, but users tend to use them as a proxy of their subjective sleep quality.

Despite the commonplace of the discrepancy between the objective sleep data and the subjective sleep experience, the mechanism of this phenomenon is poorly understood especially in the domain of the quantified-self sleep tracking. In this pilot study, we sought to provide an explanation to the discrepancy between quantified-self sleep data recorded with Fitbit and subjective sleep experience from a neuroscience perspective. We investigated the associations of the subjective sleep rating and the Fitbit measured sleep data to cortical hemodynamics in the prefrontal cortex (PFC) during the first sleep cycle. This study is the first attempt to bring a neuroscience perspective to the study of quantified-self sleep tracking. The preliminary results implied that the device measured sleep data and subjective sleep experience may characterize different aspects of the sleep process, indicating that the discrepancy between the two may not be easily addressed by simply improving device accuracy. We pointed out directions for future sleep tracking research based on our findings.

2 Related Work

Advances in wearable and mobile computing technologies have witnessed a sharp increase in the adoption of quantified-self sleep tracking technologies in recent years. These sleep tracking technologies largely fall into two categories. The first category supports users to manually log subjective sleep experience either using validated psychometric questionnaires (e.g., the Pittsburgh Sleep Quality Index; PSQI [12]) or a simple

Likert rating scale. These technologies are often made available to users in the form of a stand-alone mobile app, or as one feature in a comprehensive sleep tracking system (e.g., SleepAsAndroid, SleepBot). The other category leverages wearable or mobile sensors to objectively measure a user's sleep structure. These technologies often require users to put on a wearable tracker or to place a smartphone on bed. Popular trackers include Fitbit, Apple Watch, Mi Band and Oura Ring. Using proprietary algorithms, these devices approximate sleep structure metrics including total sleep time (TST), wake after sleep onset (WASO), sleep efficiency (SE), and sleep stages. Different devices may also have self-defined sleep metrics such as the sleep score by Fitbit, which is a weighted sum of several sleep metrics.

Studies of quantified-self sleep tracking technologies have dominantly focused on their validity/accuracy and usability. Despite of their convenience for longitudinal use in daily life settings, quantified-self sleep tracking devices may have limited accuracy depending on the type of sensors used [9]. The validity of popular sleep trackers has been well-studied in both laboratories [13] and naturalistic settings [2]. Recent findings showed that the latest models achieved reasonable accuracy for TST and SE, but not for sleep stages [2, 13]. In addition, device accuracy was found to correlate to many factors including the demographic characteristics and the sleep structure of the users [3, 21], and may also demonstrate temporal patterns [20]. The usability of quantified-self sleep trackers was also intensively studied. Previous studies have investigated how users interpret sleep data [4], users' trust towards sleep tracking technologies [7], and the challenges for these technologies to improve sleep health [9].

A common complaint related to quantified-self sleep tracking technologies is the mismatch between sleep data and users' subjective sleep experience [7]. Such mismatch is often attributed to the limited accuracy of the sleep trackers, and it is believed that improving device accuracy could help bridge the gap between the two. Nonetheless, the discrepancy between objective sleep data and subjective sleep quality has been observed even when the gold standard of sleep measurement—polysomnography (PSG)—was used. Several medical studies have demonstrated that PSG did not explain subjective sleep quality well [10]. The sleep structural metrics measured with PSG explained only 11–17% of the variance in subjective sleep quality. Discrepancy between device measured sleep structure and perceived sleep quality has been observed not only in patients with Alzheimer disease, depression, and sleep problems, but also in healthy adults [14]. For example, one study shows that healthy people who habitually slept more than 6 h at night may both overestimate and underestimate their sleep duration. In addition, the rate of overestimation significantly increased by 3 times among healthy people who slept less than 6 h at night [15]. A recent usability study into how users perceive the credibility of consumer sleep trackers also found that even when these sleep trackers agreed well to a medical reference, they not necessarily match the users' subjective sleep experience [7].

In this study, we attempted to approach the sleep misperception phenomenon from a neuroscience perspective within the context of quantified-self sleep tracking. We investigated the associations of subjective sleep rating and Fitbit measured sleep data to cortical hemodynamics in the prefrontal cortex (PFC) during the first sleep cycle. While we solely focused on Fitbit as a representative of quantified-self sleep tracking technologies, the methodology adopted in this study is readily applicable to sleep trackers of other manufacturers. This study generated new insights into the discrepancy between consumer device measured sleep data and subjective sleep experience and pointed out directions for future sleep tracking research.

3 Measuring Devices and Instruments

3.1 Wearable fNIRS System

In this study, we measured the cortical hemodynamics using a wearable functional near-infrared spectroscopy (fNIRS): the Brite 24 system developed by the Artinits Medical System Co., The Netherland. The Brite 24 measures the concentration changes of the oxyhemoglobin (ΔO2Hb) and deoxyhemoglobin (ΔHHb) in cortical brain areas. Compared to other hemodynamic imaging methods such as fMRI and PET, fNIRS is less invasive, more tolerant to motion artefacts, and allows higher temporal resolution. Compared to EEG, fNIRS has the advantage of higher spatial resolution.

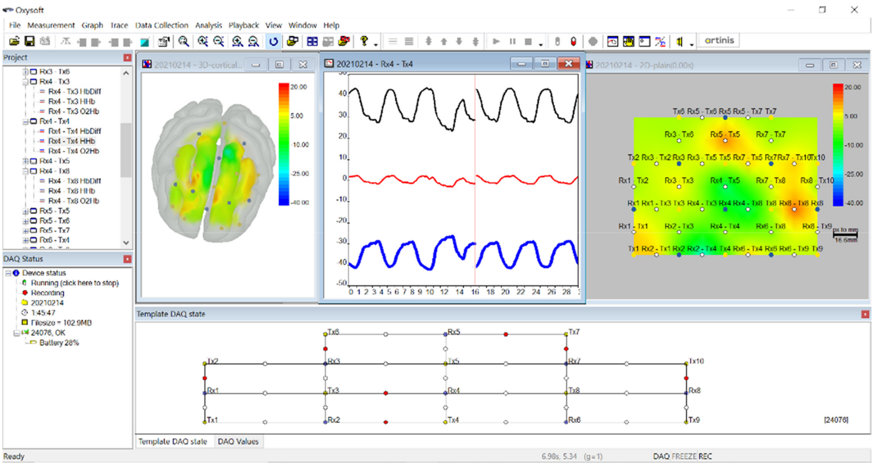


Fig. 1. A screenshot of the OxySoft companion software.

While traditional fNIRS systems often involves using bulky devices and many long cables, the Brite 24 system is a wearable and ready-to-use device that requires little set up time. The system contains 10 transmitters (Tx) that emit infrared light at the wavelengths of 760 nm and 850 nm. The infrared light travels through the sculp, the skull and the cortex. The rebounded light is captured by 8 receivers (Rx). Using the template as shown in the bottom of Fig. 1, the Txs and Rxs were configured into 27 channels, with an interoptode distance of 3 cm and penetration depth of 1.5 cm. The optodes were fixed on a soft neoprene head cap and were placed between the FpZ-F3-Cz-F4-FpZ region according to the international 10–20 EEG system. Using the neoprene head cap not only makes it easy to install and uninstall the optodes, but also ensure that the optodes were placed in the same location across different measurements. The fNIRS system has a companion software called OxySoft (Artinits Medical System Co., The Netherland). All the measurements in this study were conducted online, where the Brite 24 device was connected to the OxySoft via Bluetooth. In this way, data were regularly synchronized with the OxySoft. A screenshot of the OxySoft is illustrated in Fig. 1.

3.2 Fitbit Sense and Sleep Rating

We used Fitbit Sense—the latest model of Fitbit at the time of the study—to measure a set of sleep metrics. These metrics include total sleep time (TST), wake after sleep onset (WASO), wake ratio (WR), light sleep ratio (LR), deep sleep ratio (DR), REM sleep ratio (RR), and sleep score (SS). Participants worn the device on their non-dominant wrist and the data were automatically collected without requiring any manual input. Sleep data were synchronized with the Fitbit smartphone app the next morning upon waking up. Subjective sleep quality was collected using a 1–5 Likert scale, with 1 and 5 denoting extremely poor sleep and extremely good sleep, respectively.

4 Data Collection Protocol

In this pilot study, we adopted the N-of-1 single subject research design [16, 17]. This approach differs from the traditional cross-sectional study design in that it does not rely on the assumption of cohort homogeneity. We collected data from one healthy male participant who is Caucasian and was in his 30s. The participant did not have any diagnosed health conditions, mental problems, or sleep problems. The data collection experiment was conducted at the participant's home following a protocol listed in Table 1. The participant volunteered to shave his hair to minimize any potential interference with the signal quality of the fNIRS system.

Table 1. Data collection protocol.

Event ID	Time	Event	Device and Instrument
1	Between 22:30–23:30	Rest quietly while staying awake for 2 min	Brite 24, Fitbit Sense
2	Right after Event 1	Get in bed and lights off	Brite 24, Fitbit Sense
3	The first major wake	Take off Brite 24 and stop the measurement	/
4	Upon waking up 6:30–7:30	Rate subjective sleep quality	A 1–5 Likert scale

5 Data Analysis Protocol

The raw optical density (OD) signals recorded with Brite 24 were exported into EDF files using the companion software OxySoft. The EDF files were converted to fNIRS data type and then processed using the MNE-NIRS Python library. The OD signals of each night were trimmed between sleep start time T_S (as recorded by the Fitbit Sense) and T_S+ 90 min, which is the average length of a sleep cycle in healthy adults. The OD signal quality of each channel was analyzed using the scalp coupling index (SCI) method [18]. For each channel, the OD signals at both wavelengths were passed through a bandpass filter (0.7–1.5 Hz) to preserve only the heartbeat component. The SCI—defined as the zero-lag cross-correlation between the heartbeat component of the OD signals at both wavelengths—was then calculated and used as an indicator of the signal quality of the channel. Channels with an SCI below 0.75 were discarded. The processed OD signals were converted to ΔO2Hb and ΔHHb using the modified Beer-Lambert law (MBLL) [19], and bandpass filtered (0.02–0.18 Hz) to remove the cardiac and respiratory noise. Five time-domain features (i.e., mean, median, standard deviation, skewness, kurtosis) and three frequency-domain features (i.e., total power, peak amplitude of the frequency components, peak ratio) were derived from the ΔO2Hb and ΔHHb signals of each channel and then averaged across all channels.

Table 2. List of variables used in correlation analysis.

Category	Metric (Denotation)	Data type	Device and instrument
Cortical ΔO2Hb and ΔHHb	Mean (*mean_O2*/*mean_H*)	Continuous	Brite 24
	Median (*md_O2*/*md_H*)		
	Standard deviation (*sd_O2*/*sd_H*)		
	Skewness (*sk_O2*/*sk_H*)		
	Kurtosis (*kt_O2*/*kt_H*)		
	Total power (*tp_O2*/*tp_H*)		
	Maximum power (*mf_O2*/*mf_H*)		
	Peak ratio (*pr_O2*/*pr_H*)		
Sleep	Sleep rating (*SR*)	Ordinal	A 1–5 Likert scale
	Sleep score (*SS*)	Continuous	Fitbit sense
	Total sleep time (*TST*)		
	Wake after sleep onset (*WASO*)		
	Wake ratio (*WR*)		
	Light sleep ratio (*LR*)		
	Deep sleep ratio (*DR*)		
	REM sleep ratio (*RR*)		

The Fitbit sleep data were exported using a web application that we developed in our previous study [4]. These data were then merged with the participant's subjective sleep rating and the features derived from the hemodynamic signals by matching the date stamps. Table 2 summarizes the metrics that were used in the correlation analysis. Pearson's and Spearman's correlation coefficients were calculated pair-wisely on continuous metrics and ordinal metrics, separately. Statistical test at a significance level of $\alpha = 0.05$ was performed. Correlation coefficients were calculated using the Pandas library and statistical test was performed using the SciPy library in Python 3.8.

6 Results

The significant correlations between sleep metrics (both sleep rating and Fitbit measured sleep data) and hemodynamic features are shown in Table 3. It shows that subjective sleep rating was moderately and negatively correlated to the average *sk* and *kt* of the ΔO2Hb. No significant correlation was found between the subjective sleep rating and the features derived from ΔHHb.

Table 3. Statistically significant correlations between sleep metrics and average hemodynamic features.

Sleep metric	Feature	r	p	Sleep metric	Feature	r	p
SR	*sk_O2*	−0.516	0.049	*TST*	*md_O2*	−0.648	0.016
	kt_O2	−0.516	0.049		*mf_O2*	0.603	0.029
SS	*sd_O2*	−0.717	0.006		*md_H*	−0.59	0.034
	sd_H	−0.731	0.005	*DR*	*sd_O2*	−0.611	0.026
	sk_H	−0.589	0.034		*sd_H*	−0.594	0.032
	kt_H	−0.59	0.034		*sk_H*	−0.554	0.049
RR	*sk_O2*	−0.692	0.009	*WR*	*mean_O2*	0.735	0.004
	kt_O2	−0.716	0.006	*LR*	*sd_O2*	0.730	0.005
	tp_O2	0.673	0.012		*sk_O2*	0.726	0.005
	mf_O2	−0.653	0.015		*kt_O2*	0.698	0.008
	pr_O2	0.687	0.009		*tp_O2*	−0.538	0.058
	sd_H	−0.568	0.043		*pr_O2*	−0.599	0.031
	sk_H	−0.614	0.025		*sd_H*	0.728	0.005
	kt_H	−0.640	0.018		*sk_H*	0.745	0.003
	tp_H	0.663	0.013		*kt_H*	0.751	0.003
	mf_H	−0.695	0.008		*pr_H*	−0.702	0.007
	pr_H	0.728	0.005				

Table 4. Statistically significant correlations between subjective sleep rating and channel-wise hemodynamic features.

Rx-Tx pair	Feature	r	p
Rx3_Tx2	*md_O2*	0.514	0.050
	md_H	−0.633	0.011
Rx4_Tx3	*md_H*	0.571	0.026
	sk_H	0.561	0.030
Rx4_Tx4	*md_O2*	−0.573	0.025
	md_H	−0.531	0.042
	sk_H	0.514	0.050
Rx4_Tx5	*sk_O2*	0.691	0.004
	md_H	−0.599	0.018
Rx4_Tx8	*md_O2*	0.548	0.034
Rx5_Tx7	*sk_H*	0.620	0.014
Rx6_Tx8	*md_O2*	0.665	0.007
Rx7_Tx10	*md_O2*	0.629	0.012

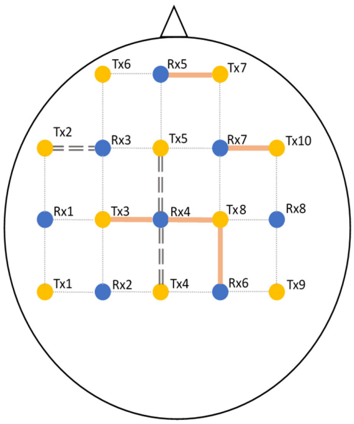
Table 5. Statistically significant correlations between Fitbit sleep score and channel-wise hemodynamic features.

Rx-Tx pair	Feature	r	p
Rx2_Tx3	*mean_H*	−0.805	<0.001
Rx2_Tx4	*mean_O2*	−0.670	0.012
	mean_H	−0.594	0.032
	md_H	−0.611	0.026
Rx3_Tx2	*mean_H*	−0.639	0.019
	md_H	−0.664	0.033
Rx3_Tx3	*mean_O2*	−0.599	0.030
	mean_H	−0.777	0.023
Rx3_Tx5	*mean_O2*	−0.852	0.007
Rx4_Tx5	*mean_H*	−0.703	0.007
Rx4_Tx8	*mean_H*	0.998	0.043
Rx5_Tx5	*mean_O2*	−0.999	0.033
Rx6_Tx9	*mean_H*	−0.740	0.023
Rx7_Tx7	*mean_H*	−0.717	0.006
Rx7_Tx8	*mean_O2*	−0.571	0.041
Rx8_Tx8	*mean_O2*	−0.926	0.003

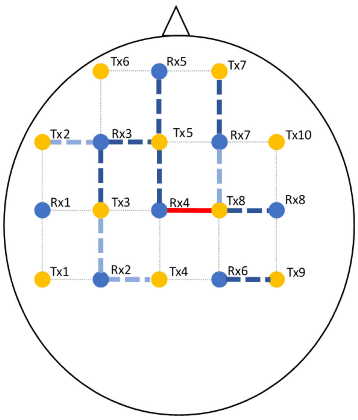
Different objective sleep metrics as measured with Fitbit Sense correlated to different hemodynamic features. Strong correlations ($|r| > 0.700$) were found between sleep score and the *sd* of ΔO2Hb and ΔHHb, wake ratio and the *mean* of ΔO2Hb, light sleep ratio and the *sd*, *sk* of ΔO2Hb and the *sd, sk, kt, pr* of ΔHHb, REM sleep ratio and the *kt* of ΔO2Hb and the *pr* of ΔHHb.

Table 4 and 5 show the statistically significant correlations between channel-wise hemodynamic features and subjective sleep rating, sleep score computed by Fitbit, respectively. Figure 2 shows a coarse visualization of the channel-wise correlations. Strong positive and negative correlation are indicated by red and dark blue connections. Moderate positive and negative correlations are indicated by orange and light blue. Grey double lines indicate both positive and negative correlations were observed. It is observed that the subjective sleep rating mainly correlated to the median of the ΔO2Hb and ΔHHb in certain channels, with positive correlation coefficients. In contrast, the sleep score computed by Fitbit mainly correlates to the mean the ΔO2Hb and ΔHHb in a different set of channels, with negative correlation coefficients.

Table 6 shows the statistically significant correlations between channel-wise hemodynamic features and other Fitbit measured sleep metrics. Figure 3, 4, 5 shows a coarse visualization of the corresponding channel-wise correlations. Broken down into sleep metrics that characterize different dimensions of the sleep structure, the total sleep time and deep sleep were mostly negatively correlated to the PFC hemodynamics, while sleep efficiency, wake after sleep onset, wake ratio and light sleep ratio (note that these features all characterize the continuity of sleep) were mostly positively correlated to the PFC hemodynamics. No correlation was observed between REM sleep ratio and the channel-wise hemodynamic features.



(a)

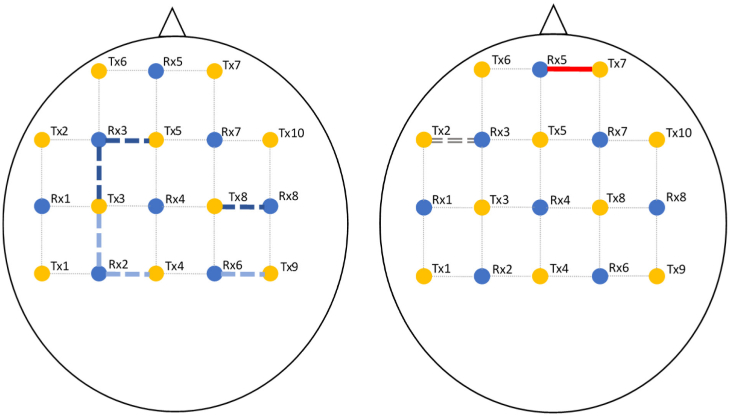


(b)

Fig. 2. A visualization of the channel-wise correlations between (a) the subjective sleep rating and the hemodynamic features, and (b) the Fitbit sleep score and the hemodynamic features. Strong positive and negative correlation are indicated by red and dark blue connections. Moderate positive and negative correlations are indicated by orange and light blue connections. Grey double lines indicate both positive and negative correlations were observed. (Color figure online)

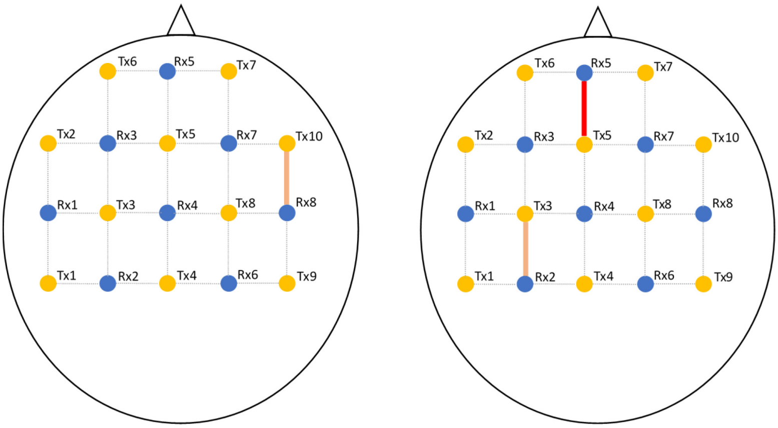
Table 6. Statistically significant correlations between Fitbit measured sleep metrics and channel-wise hemodynamic features.

Sleep metrics	Rx-Tx pair	Feature	r	p
TST	Rx2_Tx3	*mean_H*	−0.573	0.041
	Rx2_Tx4	*md_H*	−0.586	0.035
	Rx3_Tx3	*mean_O2*	−0.599	0.030
		mean_H	−0.842	0.009
	Rx3_Tx5	*mean_O2*	−0.797	0.018
	Rx6_Tx9	*mean_H*	−0.694	0.038
	Rx8_Tx8	*mean_O2*	−0.853	0.015
WASO	Rx3_Tx2	*sk_O2*	0.705	0.005
		kt_O2	−0.751	0.002
	Rx5_Tx7	*sk_O2*	0.844	<0.001
SE	Rx8_Tx10	*pr_O2*	0.642	0.018
WR	Rx2_Tx3	*mean_H*	0.677	0.011
	Rx5_Tx5	*mean_H*	0.999	0.019
LR	Rx2_Tx4	*mean_O2*	0.581	0.037
		mean_H	0.573	0.041
	Rx3_Tx2	*mean_H*	0.569	0.042
	Rx3_Tx5	*mean_O2*	0.719	0.045
	Rx6_Tx9	*mean_O2*	0.611	0.046
	Rx7_Tx8	*mean_O2*	0.566	0.044
		mean_H	0.566	0.044
	Rx8_Tx8	*mean_O2*	0.801	0.030
	Rx6_Tx4	*sk_H*	0.556	0.049
DR	Rx2_Tx4	*mean_H*	−0.641	0.018
	Rx3_Tx2	*mean_H*	−0.660	0.014
	Rx3_Tx5	*mean_O2*	−0.781	0.022
	Rx4_Tx5	*mean_H*	−0.568	0.043
		md_H	−0.572	0.041
	Rx5_Tx5	*mean_H*	−0.999	0.019
	Rx7_Tx7	*mean_H*	−0.616	0.025
	Rx7_Tx8	*mean_O2*	−0.624	0.023
		mean_H	−0.553	0.050
	Rx8_Tx8	*mean_O2*	−0.783	0.037



(a) (b)

Fig. 3. A visualization of the channel-wise correlations between (a) total sleep time and hemodynamic features, and (b) wake after sleep onset and hemodynamic features.



(a) (b)

Fig. 4. A visualization of the channel-wise correlations between (a) sleep efficiency, and (b) wake ratio and hemodynamic features.

7 Discussions

The correlation analysis results show that subjective sleep rating was moderately and negatively correlated to the average skewness and kurtosis of the ΔO2Hb, but not significantly correlated to features derived from ΔHHb. Different objective sleep metrics as measured with Fitbit correlate to different hemodynamic features. Channel-wisely, it is found that subjective sleep rating mainly correlates to the median of the ΔO2Hb and ΔHHb in a set of channels, with positive correlation coefficients. In contrast, the sleep score computed by Fitbit mainly correlates to the mean the ΔO2Hb and ΔHHb in a different set of channels, with negative correlation coefficients. This suggests that better perceived sleep quality may be positively associated to increased hemodynamics during the first sleep cycle, and the opposite is true for objective sleep metrics as measured by Fitbit. Sleep metrics related to the continuity of sleep—including WASO, SE, WR—and LR were positively associated to increased hemodynamics in the PFC during the first sleep cycle, while sleep metrics such as TST and DR were negatively associated to increased hemodynamics in the PFC during the first sleep cycle.

The analysis results imply that the discrepancy between subjective sleep experience and objective sleep data is likely due to the mismatch of the measurands. Since users' subjective consciousness and the sleep tracking device are essentially measuring different dimensions of sleep, it is reasonable to expect certain level of discrepancy between the two. As such, improving device accuracy may not help solve the discrepancy. Instead, designing future sleep tracking technologies that help users to understand the sleep misperception phenomenon could be a more plausible direction to go. It would also be interesting to explore the objective sleep metrics that correlate well to each user's subjective sleep rating, and to use them as personalized indicators of sleep quality rather than using a general set of metrics for all users like the current technologies do.

Due to the N-of-1 approach adopted in this study, the findings solely hold for this specific subject. In our future study, we plan to repeat the study protocol with a cohort of subjects to identify potential common patterns across subjects.

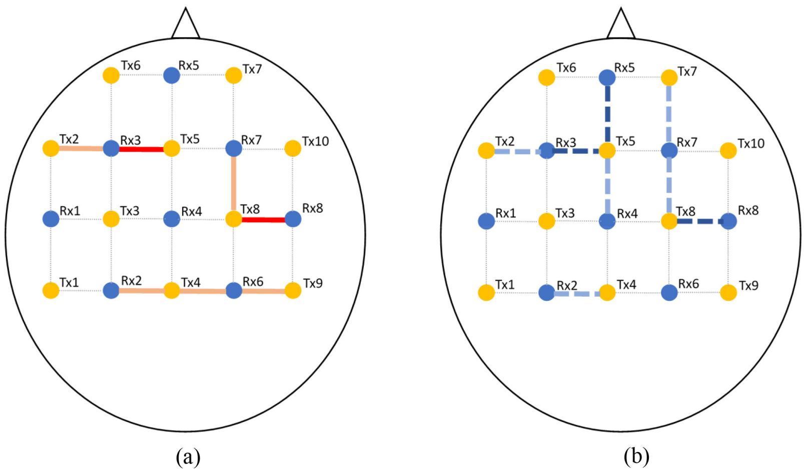


Fig. 5. A visualization of the channel-wise correlations between (a) light sleep ratio and hemodynamic features, and (b) deep sleep ratio and hemodynamic features.

Acknowledgement. This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 16H07469, 19K20141, and 21K17670.

References

1. Buysse, D.J.: Sleep health: can we define it? does it matter? Sleep **37**(1), 9–17 (2014)
2. Liang, Z., Chapa-Martell, M.A.: Validity of consumer activity wristbands and wearable EEG for measuring overall sleep parameters and sleep structure in free-living conditions. J. Healthcare Inf. Res. **2**(1–2), 152–178 (2018)
3. Liang, Z., Chapa-Martell, M.A.: Accuracy of Fitbit wristbands in measuring sleep stage transitions and the effect of user-specific factors. JMIR Mhealth Uhealth **7**(6), e13384 (2019)
4. Liang, Z., et al.: SleepExplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors. Pers. Ubiquit. Comput. **20**(6), 985–1000 (2016). https://doi.org/10.1007/s00779-016-0960-6
5. Weatherall, J., et al.: Sleep tracking and exercise in patients with type 2 diabetes mellitus (step-D): pilot study to determine correlations between Fitbit data and patient-reported outcomes. JMIR Mhealth Uhealth **6**(6), e131 (2018)
6. Bian, J., et al.: Exploring the association between self-reported asthma impact and Fitbit-derived sleep quality and physical activity measures in adolescents. JMIR mHealth and uHealth **5**(7), e105 (2017)
7. Liang, Z., Ploderer, B.: How does Fitbit measure brainwaves: a qualitative study into the credibility of sleep-tracking technologies. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **4**(1), Article 17 (2020)
8. Liu, W., Ploderer, B., Hoang, T.: In bed with technology: challenges and opportunities for sleep tracking. In: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction, pp. 142–151. Parkville, VIC, Australia (2015)
9. Liang, Z., Ploderer, B.: Sleep tracking in the real world: a qualitative study into barriers for improving sleep. In: Proceedings of the 28th Australian Conference on Computer-Human Interaction, pp. 537–541. Launceston, Tasmania, Australia (2016)
10. Kaplan, K.A., et al.: When a gold standard isn't so golden: lack of prediction of subjective sleep quality from sleep polysomnography. Biol. Psychol. **123**, 37–46 (2017)
11. Baker, F.C., Maloney, S., Driver, H.S.: A comparison of subjective estimates of subjective estimates of sleep with objective polysomnographic data in healthy men and women. J. Psychosom. Res. **47**(4), 335–341 (1999)
12. Buysse, D., et al.: The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res. **28**(2), 193–213 (1989)
13. De Zambotti, M., et al.: A validation study of Fitbit Charge 2 compared with polysomnography in adults. Chronobiol. Int. **35**(4), 465–476 (2017)
14. Hsiao, F.-C., et al.: The neurophysiological basis of the discrepancy between objective and subjective sleep during the sleep onset period: an EEG-fMRI study. Sleep **41**(6), zsy056 (2018)
15. Fernandez-Mendoza, J., et al.: Sleep misperception and chronic insomnia in the general population: the role of objective sleep duration and psychological profiles. Psychosom. Med. **73**(1), 88–97 (2011)
16. De Groot, M., et al.: Single subject (N-of-1) research design, data processing, and personal science. Methods Inf. Med. **6** (2017)
17. Shaffer, J.A., et al.: N-of-1 randomized intervention trials in health psychology: a systematic review and methodology critique. Ann. Behav. Med. **52**, 731–742 (2018)
18. Pollonini, L., et al.: Auditory cortex activation to natural speech and simulated cochlear implant speech measured with functional near-infrared spectroscopy. Hear. Res. **309**, 84–93 (2014)
19. Delpy, D.T., et al.: Estimation of optical pathlength through tissue from direct time of flight measurement. Phys. Med. Biol. **33**, 1433–1442 (1988)

20. Liang, Z., Chapa-Martell, M.A.: Combining numerical and visual approaches in validating sleep data quality of consumer wearable wristbands. In: Proceedings of IEEE PerCom Workshops (IQ2S Workshop), pp.777–782. Kyoto, Japan (2019)
21. Liang, Z., Chapa-Martell, M.A.: A multi-level classification approach for sleep stage prediction with processed data derived from consumer wearable activity trackers. Front. Dig. Health **3**, 665946 (2021)