Is Fitbit Fit for Sleep-Tracking? Sources of Measurement Errors and Proposed Countermeasures

ID	Age	Device Type
G1	20s	Charge HR
G2	30s	Charge HR
G3	40s	Charge HR
G4	30s	Flex
G5	50s	Charge HR
G6	40s	Charge HR
G7	50s	Flex
P8	30s	Charge HR
P9	30s	Charge
P10	30s	Flex
P11	40s	Charge HR
P12	40s	Flex
G13	20s	Charge 2
G14	30s	Charge 2

Table 1: Demographic information of the participants. The first alphabet of ID number indicates sleep quality measured by Pittsburgh Sleep Quality Index [18] (G=good sleeper, P=poor sleeper).

Zilu Liang
The University of Tokyo, Japan
z-liang@t-adm.t.u-tokyo.ac.jp

Bernd Ploderer
Queensland University of Technology, Australia
b.ploderer@qut.edu.au

Mario Alberto Chapa-Martell
CAC Corporation, Japan
mchapam0300@gmail.com

PervasiveHealth '17, May 23–26, 2017, Barcelona, Spain

ACM ISBN 978-1-4503-6363-1/17/05…$15.00
https://doi.org/10.1145/3154862.3154897

Abstract
It is now easy to track one's sleep through consumer wearable devices like Fitbit from the comfort of one's home. However, compared to clinical measures, the data generated by such consumer devices is limited in its accuracy. The aim of this paper is to explore how users perceive accuracy issues, possible measurement errors and what can be done to address these issues. Through an interview study with 14 Fitbit users we identified three main sources of errors: (1) lack of definition of sleep metrics, (2) limitations in underlying data collection and processing mechanisms, and (3) lack of rigor in tracking approach. This paper proposes countermeasures to address these issues, both from the aspect of technological advancement and through engaging end-users more closely with their data.

Author Keywords
Sleep; health; personal informatics; wearable; Fitbit; data quality; HCI.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

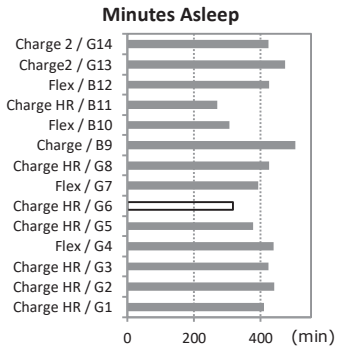

Figure 1: Average minutes asleep measured by Fitbit. The notion on x-axis indicates device type / participant ID. White bars indicate unintuitive results, e.g. participant G6 considered herself as good sleeper, but her data showed very short sleep duration.

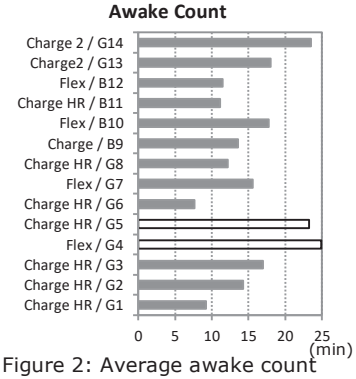

Figure 2: Average awake count measured by Fitbit. Participant G4 and G5 considered themselves as good sleepers, but their data showed many awakenings.

Introduction

Wearable technologies such as Fitbit offer the benefit of easy and affordable sleep-tracking from the comfort of one's home. Unlike clinical studies which only take snapshots of one's sleep over a single night, consumer devices support the collection of data over extended periods of time. These data can generate various benefits, e.g., to help healthy people understand individual's sleep patterns, or to detect possible sleep problems like insomnia [1].

A major shortcoming of consumer sleep-tracking devices is the lack of accuracy of the data collected. Experts have warned users of the risk of developing a flawed or incomplete picture of their sleep status based on inaccurate measurements using Fitbit [2, 6], and user concerns about the accuracy of fitness trackers in general have been raised [3]. However, a close examination of how end-users perceive and respond to accuracy and measurement errors is missing.

In this paper, we delve deeper into the data accuracy issue and set out to understand the sources of the measurement errors of Fitbit devices, and we propose design recommendations to enhance the accuracy of Fitbit sleep-tracking. The errors that we focus on are the wrong measures that deviated significantly from the truth values. Through an interview study with 14 Fitbit users who had been tracking their sleep, we found that the measurement errors with Fitbit were not only a technical issue but also related to human factors. We found three main sources of measurement errors that harm the usefulness of self-tracking sleep data: (1) lack of definition of sleep metrics, (2) limitation in 4iand (3) lack of rigor in collecting data. To address these issues, we propose three design recommendations from the aspects of both the technology and the end-users in the discussion section. In summary, we hope that findings and recommendations presented in this paper can benefit not only designers and developers of home sleep-tracking technologies who want to make better devices, but also researchers and clinical practitioners who seek to gain accurate data through Fitbit.

Method

We interviewed 14 Fitbit users (11 females and 3 males) in Melbourne and Tokyo to discuss their sleep-tracking practices and to investigate possible sources of measurement errors. They came from a variety of professions. The duration of sleep tracking in this study was 36±14 days. Participant G13 and G14 also tracked their subjective sleep quality using a sleep diary. The demographic information of the participants is summarized in Table 1. We audio-recorded and transcribed all interviews and the qualitative data were analyzed using iterative affinity analysis.

Findings

The basic statistics of the sleep data collected using Fitbit devices is shown in Figure 1~4. Some of the results are unintuitive (represented by white bars), indicating potential measurement errors. Through affinity analysis we identified the following three main sources of measurement error in Fitbit.

Lack of Definition of Sleep Metrics

Fitbit infers sleep quality from measurable physiological signals, i.e. through movements of the wrist. However, it is not clear to users as to which kind of movement is getting measured and how the movements are translated into sleep status. "*Twenty-two restlessness;*

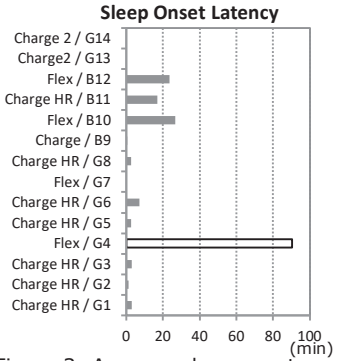

Figure 3: Average sleep onset latency measured by Fitbit. Participant G4 considered herself as good sleeper, but her data showed long sleep onset latency.

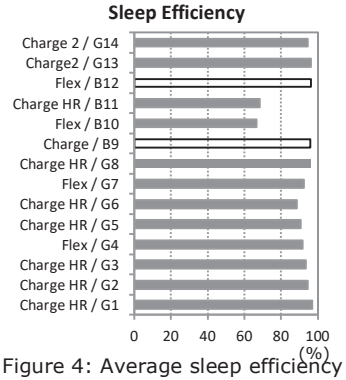

Figure 4: Average sleep efficiency measured by Fitbit. Participant B9 and B12 considered themselves as bad sleepers, but their data showed that their sleep efficiency was satisfying.

that's a lot of movements. But what kind of movement we are talking? Arm movement or the whole body movement?" (G3). Although Fitbit does provide some information on how some sleep metrics were measured, the description is ambiguous. Six out of 14 participants mentioned that the border line between "awakenings" and "restlessness" is obscure. *"I think it would be helpful if Fitbit differentiate awakenings and restlessness. Because when I'm awake, I know it. But restlessness is like additional information that I'm not aware of"* (G6). As is shown in Figure 5, Fitbit's measurement on the number of awakenings/restless significantly deviated from the subjective sleep experience of participant G14.

Limitation on Underlying Sensors and Algorithms

The second source of measurement errors in Fitbit measurements originates from the limitation of underlying mechanism of the embedded sensors and algorithms processing these data. Fitbit devices dominantly rely on accelerometer data to measure sleep. Fitbit devices tend to falsely record sleep when a user is awake but not moving, e.g. reading in bed or watching TV on sofa. This limitation of Fitbit is well-known to users, as many users noticed the wrong measurements when they compare Fitbit data against their subjective sleep experience. "*Sometimes when I was in bed reading, fitbit counted that I was in sleep so you see here: I slept for 10 hours, which was not true"* (B11). In spite of Fitbit's continuous effort to solve the false-sleep detection problem, the overcorrected algorithms in the latest model of Fitbit tend to aggressively record awake when users have actually fallen asleep. Participant G14 mentioned that his Fitbit Charge 2 eliminated the sleep segment between the time he fell asleep while reading books in bed and his first awakenings. Another limitation of the underlying mechanism of Fitbit was that it may capture all the movements that it detected, regardless of whether the movements were initiated by the user who was wearing the device or were actually from the environment. "*I don't know that was the movement from my husband or from mine"* (B10).

Lack of Control in Home Sleep Tracking

Self-tracking in free conditions is usually conducted in an uncontrolled manner. Some measurement errors may be attributed to the usage patterns. For Fitbit models that require manual switch in and out of sleep mode, measurement errors may be produced if a user forgot to tap the device when they entered or got out of bed. "*This is a Flex; you have to tap it to stop the sleep tracking mode. That just means I was awake in the morning"* (B10). This echoes the quantitative data shown in Figure 6: the average sleep onset latency measured by Fitbit Flex is generally longer than that measured by Fitbit HR Charge.

Another common source of measurement error is that some users do not wear their devices regularly. Fitbit was developed under the assumption of continuous use. However, users may fail to do so either intentionally or subconsciously. Discontinued use, especially before entering bed or after getting out of bed, could lead to wrong measurements. *"One day I woke up around 7:30. I took off my Fitbit and left it on the bed. Later when I synched my Fitbit I noticed the data showed I didn't wake up until 10:00"* (G13).

We guided the participants to reflect on their usage patterns and how such patterns may impact the measurement accuracy of Fitbit. One participant

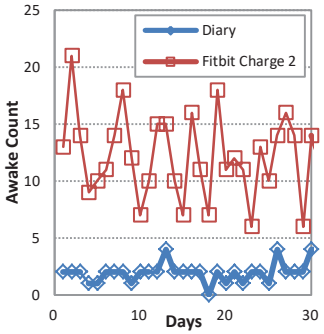

Figure 5: Large disparity was identified between Fitbit measurements and subjective sleep experience on awakening count (participant G14).

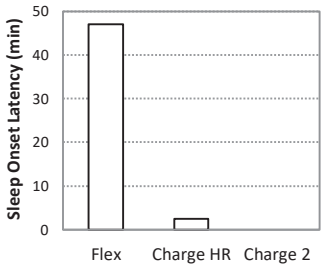

Figure 6: Average sleep onset latency measured by Flex, Charge HR, and Charge 2. The measurements by Flex are significantly longer than that by Charge HR and Charge 2, indicating potential measurement errors.

mentioned that it raised his awareness that some of the measurement errors were actually attributed to his own behavior. *"Sometimes it's difficult to be aware that you are doing something that is affecting the readings the device gives you"* (G14).

Discussions

Based on the above findings, we propose three countermeasures for addressing these errors: (1) provide clear and consistent definition on sleep metrics and how they are measured, (2) encourage collaborations and seek validation from end-users, and (3) enhance data quality using advanced data analysis techniques. Due to page limit, we only briefly discuss each countermeasure below. Previous actigraph validation research found that the agreement between actigraphy and PSG varies depending on the definition of sleep metrics [4]. Commercial sleep trackers should properly define the sleep metrics that are measured and ensure that the definitions are consistent with those in sleep research. Secondly, devices may be designed to encourage collaborations from users by providing adaptive compliance feedback to remind the users gently on how and when to wear the device. Another way is to engage users to validate their sleep records by allowing greater freedom for data editing. From a technical aspect, we may design new algorithms to cancel out measurement errors based on insights on the error distribution of Fitbit, and we may also apply data fusion technology to integrate information from other sources [5].

Conclusion

This study offers a rich description of how sleep-trackers experience measurement errors, and the ways in which these errors diminish the accuracy and usefulness of their sleep data. The main sources of measurement errors are due to a lack of definition of sleep metrics, limitations in the underlying ways of sensing and processing data, due to a lack of rigor in tracking data. To address these problems, this paper articulates countermeasures, i.e., novel ways of defining sleep, closer engagement between sleep-trackers and their data, and advanced data analysis techniques. We hope that these suggestions provide clear directions for future research and better insights for self-trackers

References

1. Shelgikar, A. V., Anderson, P. F., Stephens, M. R. Sleep tracking, wearable technology, and opportunities for research and clinical care. *CHEST 150*, 3 (2015), 732-743.
2. Kolla, B. P., Mansukhani, S., and Mansukhani, M. P. Consumer sleep tracking devices: a review of mechanisms, validity and utility. Expert Review of Medical Devices 13, 5 (2016), 497-506.
3. Liang, Z., Ploderer, B. Sleep tracking in the real world: a qualitative study into barriers for improving sleep. In Proceedings of OzCHI 2016, 537-541.
4. Sadeh, A., Hauri, P., Kripke, D., Lavie, P. The role of actigraphy in the evaluation of sleep disorders. *SLEEP 18*, 4 (1995), 288-302.
5. Goverdovsky, V., Looney, D., Kidmose, P., Papavassiliou, C., and Mandic, D. P. Co-located multimodal sensing: a next generation solution for wearable health. *IEEE Sensor Journal 15*, 1 (2015), 138-144.
6. Yang, R., Shin, E., Newman, M. W., & Ackerman, M. S. (2015). When fitness trackers don't "fit": end-user difficulties in the assessment of personal tracking device accuracy. In Proceedings of UbiComp 2015, 401-410.