Digital Phenotype Collection System Utilizing Smart Devices

Jeong Mook Lim
Electronics and Telecommunications Research Institute
Daejeon, South Korea
jmlim21@etri.re.kr

Seungeun Chung
Electronics and Telecommunications Research Institute
Daejeon, South Korea
schung@etri.re.kr

Se Won Oh
Electronics and Telecommunications Research Institute
Daejeon, South Korea
sewonoh@etri.re.kr

Kyoung Ju Noh
Electronics and Telecommunications Research Institute
Daejeon, South Korea
kjnoh@etri.re.kr

Hyuntae Jeong
Electronics and Telecommunications Research Institute
Daejeon, South Korea
htjeong@etri.re.kr

***Abstract*— In this paper, we introduced a system that collects daily sleep data and daily routine data to explore the relationship between sleep and daily life. The daily data is quantified into various social and behavioral fingerprint information by analyzing sensor data obtained from carrying and using smartphones and smartwatches. By analyzing the quantified daily data in combination with the quality of sleep on the same day, we can discover factors that greatly affect sleep during daily life. To analyze the association between daily life and sleep quality, we described the sensor data collected daily and the system that generates descriptive features processed from the sensor data.**

Keywords*—*sleepcare, digital phenotype, smartphone, smartwatch

I. INTRODUCTION

In the modern world, smartphones and smartwatches are rapidly becoming ubiquitous and having a significant impact on individuals' daily lives. Smartphones are becoming increasingly popular around the world, with the majority of the population in major countries owning and using a smartphone[1, 2]. Their portability and versatility make them an essential tool for everyday life, and smartphone penetration continues to grow as technological advances and economic factors make them relatively affordable. Smartwatches were relatively slow to take off, but their popularity is growing due to technological advancements and design improvements. In particular, smartwatches have attracted the attention of users by characterizing health monitoring and healthcare functions, and for some users, they are used in conjunction with smartphones to complement some aspects of their lives.

Given that smartphones and smartwatches have rapidly become ubiquitous in modern life, and that these social and behavioral digital fingerprints can be collected in near realtime, they can be leveraged to reveal personal behavioral characteristics and lifestyle habits of which users may be unaware. According to recent studies, the information that can be collected from smartphones during daily life includes spatial trajectory information using GPS sensor data, behavioral and movement information using accelerometer sensor data, socio-dynamic information using call or social app usage data, and environmental information about the user's space using acoustic and illumination data[3]. Given the advances in smartphone sensor technology, the level of information collected is unlikely to be limited to the current level. Smartwatches can also collect a variety of physical activity data, such as heart rate, blood pressure, steps, and sleep patterns through bio-signal analysis. By monitoring a user's physical activity and lifestyle patterns during daily life, it is possible to systematically record and track a user's overall routine. This identification of individual characteristics will enable the provision of personalized services to individuals in a variety of service areas, including online learning and education services, personalized marketing services, content recommendation services, smart home services, and personal transportation services.

In particular, real-time, quantified data collected from a user's digital devices is defined as a digital phenotype, which is a representation of an individual in the healthcare field [4]. These digital phenotypes provide valuable insights into an individual's health status and lifestyle habits, and are used to develop systematic methods for personalized healthcare, such as early detection of various diseases and tracking their progression [5].

Sleep is an important factor in maintaining physical and mental health. Not getting enough sleep can lead to a variety of health problems, including fatigue, poor concentration, cognitive impairment, cardiovascular disease, obesity, diabetes, depression, and anxiety [6, 7]. Traditional sleep management methods include surveys, polysomnography, and sleep diaries, but these methods are expensive and cumbersome to use. Smartphones and smartwatches can be effectively used for sleep management and prevention of sleep diseases because they can measure various vital signs, behavioral information, and environmental information just by using them.

In this paper, we present a daily life data collection system that utilizes smartphones and smartwatches to identify key digital phenotypes related to sleep quality during daily life. We are currently utilizing this system to accumulate daily data and standard sleep metric from physically healthy adults with self-reported sleep disorders. The daily data and standard sleep indicator data, together with standard sleep surveys collected through periodic consultations with sleep specialists and diagnostic results from specialists, constitute a dataset for research on sleep-related AI technologies.

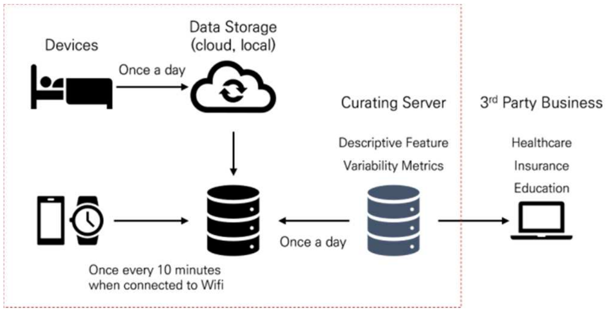

Figure 1. Sleepcare Platform

The results of this study will be applied to the development of a sleep management system that will allow users to self-identify factors in their daily lives that affect their sleep status and guide behavioral changes in their daily lives. It is expected that users will be able to improve their own sleep quality and prevent fatigue and stress caused by sleep disorders, thus contributing to the maintenance of an ongoing healthy life.

II. Sleepcare Platform

The sleep management platform is organized as shown in Figure 1. It consists of a daily data collection system using smartphones and smartwatches, a storage system that stores the collected data, an information extraction system that quantifies daily life and finds meaning from the stored raw data, and a daily sleep analysis system that analyzes the relationship between the raw data or quantified information and the sleep indicators of the day. Of these, this paper introduces the data collection system and storage system.

- Data collection systems

Data collected by apps on smartphones consists of two types of data: passive data, which is sensor data collected periodically without user intervention, and active data, which is collected with user participation, such as surveys. Active data includes surveys that users complete twice daily, once after waking up and once before going to bed. [Figure 2] The content of each questionnaire is based on items that can affect sleep, such as physical fatigue, mental fatigue, mood, and stress levels, while exercise, caffeine, and alcohol consumption are collected before bedtime, and dream-related information is collected after waking up.

For passive data, you decide which of the sensors provided by your device to use based on experimentation. Use GPS sensor data for information about the space the user is in or movement, and use the accelerometer sensor for movement information such as posture or exercise. For these sensors, battery consumption is a concern, so it's important to determine the optimal sampling rate through internal testing. For indoor spaces where GPS signals are not available, information can be collected from nearby Wi-Fi or Bluetooth devices to determine movement or location in the room. Smartphone usage information is a potential indicator of social activity.

To infer how actively an experimenter participates in online social activities, we collect information about the time and frequency of use of applications in the social category of the smartphone. We can also collect information about the on/off of the smartphone screen by time of day and the frequency of use of apps outside the social category to analyze online activities from different perspectives. For background information, we use ambient noise and light sensor values.

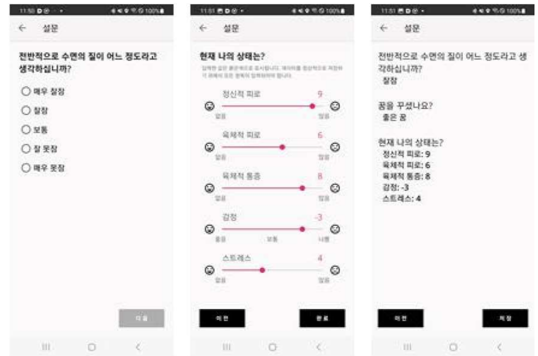

Figure 2. Morning Survey Screen (in Korean)

Table 1. Summary of Collected Data

Data	Frequency (Hz)
Accelerometer	50
GPS	0.2
Activity	1/60
Light	1/600
Wifi	1/600
Bluetooth	1/600
Background Noise	1/120
Heart Rate(W)	1/60
Steps(W)	1/60
Light(W)	1/600

Ambient noise data can be used to infer the subject's emotional state and the space they are in. Background noise information provides rich background information about the subject's space and the subject itself, but due to privacy concerns, we store probability values of audio ontology labels created using the YAMnet package [8] rather than actual audio clips.

Illumination information stores values obtained from the illumination sensors of smartphones and smartwatches. In addition, the smartwatch is used to collect heart rate and step counts.

Currently, the data collected for sleep care service includes acceleration, GPS, activity, Wifi, Bluetooth, app usage, background noise, and light data from the smartphone(Android 13), and heart rate, light, and step data from the watch(wearOS 3)[Table 1].

The standard sleep data is collected from the Withings sleep tracking mat [9], the data is sent to the Withings' cloud server, and the data management system periodically saves the standard sleep data stored on the cloud server to the data management system, including sleep score, bedtime, wake-up time, time spent in each sleep stage, total sleep time, snoring count, WASO(Wakeup time After Sleep Onset), etc.

- Data management systems

The data management system provides the ability to create profiles of collected data, manage experiment participants, and manage raw data. The experiment administrators assigns a unique ID and password to each experiment participant and stores the participant's data profile and information such as gender, age, and dominant hand.

The system flexibly manages sensor types and data sampling rate for smartphones and smartwatches based on different experiment settings. Experiment administrators can select a pre-defined data profile to enroll participants in the experiment, and when participants log in to the system, they can set the required sensor data and sampling frequency for their smartphones. The raw data collected is stored on the smartphones and smartwatches, and when connected to Wi-Fi, the system sends the data to a database for storage. This facilitates the management and maintenance of the experimental data.

- Descriptive features extraction system

The descriptive information extraction system provides the ability to quantify a user's day by analyzing the raw data stored in the data management system. The raw data stored in the data management system is converted into information about activities, smartphone usage, transportation, user experience, and location by aggregating the data at daily time intervals.

Activity information uses two types of sensor data. First, smartphone acceleration data is used to estimate representative behaviors for each minute, and then chunks are organized into 10-minute chunks to record the number of activity changes for each chunk, the time of active behavior such as walking and running, and the time of vehicle use. Second, we use heart rate data from smartwatches to extract the average heart rate, standard deviation, and exercise intensity of 10-minute chunks. Exercise intensity was mapped into rest (~50%), moderate (50-63%), active (64-75%), and vigorous (76%~) stages based on the maximum heart rate considering the subject's age, and expressed as a cumulative weighted sum (weights: 1, 2, 3, 4).

Smartphone usage information was based on application usage information provided by the smartphone OS, mapped to used applications and category information from the App store, and the usage time for each category was aggregated. Transportation mode was categorized into stationary, walking/running, bicycle, and car using GPS sensor data. When using a bicycle or car, we recorded the speed and distance traveled per chunk. Information about the user's environment is stored as illumination information and Google Sound Vocabulary information using ambient noise. The illumination data is classified into brightness categories (dark, home bright, office bright, outdoor bright) by time chunk, and the ambient noise data collected by time chunk is converted into probability values of Google Sound labels and the top 10 labels are stored [10]. Place information was assigned a stay point ID by calculating the space where the user stays for a certain period of time for each chunk using GPS sensor data, and information about spatial movement in the building was extracted using Wifi and Bluetooth data. The descriptive features extracted from the raw data are analyzed in combination with standard sleep data collected daily, such as sleep quality, sleep score, total sleep time, snoring frequency, etc. to identify the main factors affecting sleep quality during each user's daily routine.

III. CONCLUSION

In this paper, we introduced a sleep management platform for collecting data to manage sleep quality through daily behavioral monitoring. We showed a system that collects and analyzes information about the user's behavior and environment from smart devices that accompany the user. The results of this research will be applied to the development of a sleep management system, which will be used to help users identify factors in their daily lives that affect their sleep status and induce behavioral changes in their daily lives.

ACKNOWLEDGMENT

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [23ZS1100, Core Technology Research for Self-Improving Integrated Artificial Intelligence System].

REFERENCES

[1] Mobile Fact Sheet Pew Research Center. 2019. https://www.pewresearch.org/internet/fact-sheet/mobile/

[2] Zippia. "20 Vital Smartphone Usage Statistics [2023]: Facts, Data, and Trends On Mobile Use In The U.S." Zippia.com. Apr. 3, 2023, https://www.zippia.com/advice/smartphone-usage-statistics/

[3] Torous, J., Kiang, M. V., Lorme, J. & Onnela, J. P. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. Jmir Mental Health 3, e16 (2016)

[4] Onnela J-P, Rauch SL (2016) Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 41(7):1691–1696

[5] Torous J, Onnela J, Keshavan M (2017) New dimensions and new tools to realize the potential of RDoC: digital phenotyping via smartphones and connected devices. Trans Psych 7(3):1053–1053

[6] O'Brien EM, Mindell JA. Sleep and risk-taking behavior in adolescents. Behav Sleep Med 2005;3:113-133

[7] Smaldone A, Honig JC, Byrne MW. Sleepless in America: inadequate sleep and relationships to health and well-being of our nation's children. Pediatrics 2007;119:S29-S37

[8] YAMNet,https://github.com/tensorflow/models/tree/master/research/audioset/yamnet

[9] Withings Sleep Tracking Mat, https://www.withings.com/kr/en/sleep

[10] Google Audio Ontology, http://research.google.com/audioset/ontology/index.html