# Smartwatch Wearing Behavior Analysis: A Longitudinal Study

HAYEON JEONG, KAIST
HEEPYUNG KIM, KAIST
RIHUN KIM, KAIST
UICHIN LEE, KAIST
YONG JEONG, KAIST

Smartwaches are the representative wearable or body-worn devices that provide convenient and easy information access. There is a growing body of research work on enabling novel interaction techniques and understanding user experiences of smartwatches. However, there is still lack of user experience research on wearing behaviors of smartwatches, which is critical for wearable device and service design. In this work, we investigate how college students wear smartwatches and what factors affect wearing behaviors by analyzing a longitudinal activity dataset collected from 50 smartwatch users for 203 days. Our results show that there are several temporal usage patterns and distinct groups of usage patterns. The factors affecting wearing behaviors are contextual, nuanced, and multifaceted. Our findings provide diverse design implications for improving wearability of smartwatches and leveraging smartwatches for behavioral changes.

CCS Concepts: •**Human-centered computing →Ubiquitous and mobile devices; Empirical studies in ubiquitous and mobile computing;** *Empirical studies in HCI;*

General Terms: Design, Algorithms, Performance

Additional Key Words and Phrases: Smartwatch, User Experience, Wearable Device, Activity Tracker, Personal Informatics

**ACM Reference format:**
Hayeon Jeong, Heepyung Kim, Rihun Kim, Uichin Lee, and Yong Jeong. 2017. Smartwatch Wearing Behavior Analysis: A Longitudinal Study. *PACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 60 (September 2017), 31 pages.
DOI: http://doi.org/10.1145/3131892



## 1 INTRODUCTION

A smartwatch is a wearable device that has gained significant traction in recent years. A smartwatch has the ability to aid users' daily lives because it complements their smartphones by providing convenient and easy information access such as checking notifications promptly. In addition, human activity sensing capabilities with smartwatches will enable a host of mobile health applications as promised by recent mobile health platforms such as Google Fit, Apple HealthBook, and Microsoft Healthvault [20, 44]. Many smartwatches have been released in recent years, not only by major smartphone manufacturers such as Apple and Samsung but also by startups such as Pebble. Furthermore, there is a growing body of research work on enabling novel interaction techniques [5, 20, 38, 54] and understanding user experiences of smartwatches [9, 32, 42, 48].

In particular, prior studies examined various user experience topics such as usage motives, usability issues, and usage patterns. For example, researchers identified that smartwatches are mainly used for checking notifications


2474-9567/2017/9-ART60 $15.00
DOI: http://doi.org/10.1145/3131892

and tracking physical activities [32, 42, 48], and micro-interactions with smartwatches (i.e., short and frequent access) are considered very useful in various contexts (e.g., checking notifications while socializing) [9, 42, 48].

The major departure from conventional mobile devices such as smartphones and tablets is that smartwatches are wearable or body-worn devices. Users can interact with conventional mobile devices, as long as they are nearby. However, wearable devices, such as smartwatches, are deemed useful if users wear them; otherwise, they cannot fully reap the benefits of smartwatches (e.g., interacting with a glanceable display for prompt checking). In general, understanding wearing behaviors is critical for designing wearable devices, and thus, it is important to investigate how people wear smartwatches and to examine what factors such as usability [33, 37, 42] and aesthetic concerns [10, 11, 19] affect wearing behaviors. Furthermore, smartwatches can be considered as a representative wearable device. Overall, understanding wearing behavior of smartwatches would provide various insights into designing other wearable devices such as activity trackers and smart glasses.

However, there is still lack of user experience research on wearing behaviors of smartwatches. Prior studies mostly examined the motives and usability issues [9, 42, 48], and measurement studies were to limited to small-scale, short-term behavior analyses on micro-interactions [36]. None of the prior studies attempted to characterize wearing behaviors of smartwatches, which requires a long-term, in-vivo measurement study due to unique characteristics of smartwatches, i.e., ubiquity of access, micro-interaction styles, and mobility of users.

In this study, we aim to answer the following research questions about smartwatch wearing behaviors: (1) what are the temporal wearing behaviors, and what motivates these patterns? (2) What are the major factors that affect wearing and taking off behaviors? Towards this goal, we performed a longitudinal user study by distributing Apple Watches to 50 college students and collecting activity tracking data (i.e., step counts and heart rates) for just over 200 days. We chose the Apple Watch because it currently has the largest market share in the world (45.6% as of Q3 2016) [8]. Our work investigated the college students as a focusing lens on emerging wearable technologies as in prior campus health studies [43, 52]. College students are of special interests to the research communities because of their developmental dynamics and relative independence from social roles and expectations [21, 43, 52].

This quantitative study gave us an understanding of how frequently, how long, and when users wear their watches in a day. In addition, we conducted a survey and interview study with the participants who provided their activity tracking data. Through this qualitative study, we obtained detailed reasons as to why such behaviors were observed and what the major factors that affect wearing and taking off behaviors are. The major contributions of this paper are the following:

- This work is the first longitudinal measurement study about wearing behaviors of smartwatches. We used a mixed method by quantitatively analyzing temporal wearing patterns and qualitatively investigating reasons and factors related to wearing behaviors.
- We built an activity data collection platform for the Apple Watch and collected a longitudinal dataset. Using the collected dataset, we then proposed a novel approach that infers whether a user wears a watch or not using step count and heart rate data. We argue that conventional health trackers mostly report step counts and heart rates, but they cannot be directly used to infer the wearing state because the heart rate input is less stable with mobility, and step counts may be tracked even when the tracker is not worn. Our machine learning model can achieve over 95% accuracy with a detection time granularity of 8 minutes in everyday scenarios.
- Our off-line data analysis showed that the average daily wearing times were 10.99 hours and 8.40 hours on weekdays and weekends, respectively. The average number of take-off events per day was estimated to be 3.17. We found that subjective reports were less accurate: users overestimated the average wearing duration as 13.38 hours, whereas they underestimated the average number of take-off events per day as 2.06. Our cluster analysis identified three distinct groups of usage patterns, namely work-hour wearers, active-hour wearers, and all-day wearers. Temporal usage analysis showed that usage density of a user is

fairly high and consistent with its average of 88.7%, and a majority of users (92%) tended to only have short break duration (with the median break length of 5 days or less).

- We present the major themes related to wearability that were uncovered from content analysis of the interview results, such as micro-interaction needs in daily routines, smartwatch charging, aesthetic concerns, and activity/exercise tracking accuracy. Based on the findings from this study, we suggest several design implications for improving wearability of smartwatches and leveraging smartwatches for behavioral changes.

Note that our study has several limitations about generalizability of the findings due to limited data collection. To generalize our findings, there should be comparative studies with other smartwatch devices and diverse user groups.

The rest of the paper is organized as follows. In Section 2, we review related work in this field by examining prior studies about user experiences of smartwatches and activity trackers. In Section 3, we present a quantitative study of wearing behaviors. After providing a brief overview of the overall data processing, we detail the data collection methods and illustrate our wearing state classification models. We then provide the analysis results about temporal patterns of wearing behaviors, i.e., daily wearing patterns, diurnal wearing patterns, and take-off event patterns. In Section 4, we aim to corroborate our quantitative findings in Section 3 with a qualitative study. We look at the reasons behind our quantitative findings by investigating a variety of contextual elements such as places, activities, and social circumstances. We performed an online survey and in-depth interviews and conducted content analysis to corroborate our major findings about our research questions from earlier sections. In Section 5, we summarize our major findings, discuss how our findings are related to the prior studies, and examine the limitations of our work. In Section 6, we conclude our work and provide possible directions for future work.

## 2 RELATED WORK

We provide an overview of prior studies on user experiences of smartwatches and fitness trackers.

### 2.1 Smartwatch Using Behaviors

Prior studies on smartwatches mostly focused on developing new interaction methods and creating new smartwatch services [5, 20, 38, 54]. In recent years, there have been a growing body of user experience research on smartwatches, e.g., how people use their smartwatches in their daily lives, whether they are usable, and what users' needs are. Our work complements the prior studies on user experience research by investigating longitudinal wearing behaviors with the activity tracker dataset from 50 Apple Watch users for just over 200 days.

Due to low smartwatch adoption, researchers have used various methods to investigate user experiences. Lyons [32] conducted a user study by recruiting 50 smartphone users who were using dumb digital watches. To find implications for smartwatch design, he asked the participants how they had been using their conventional digital watches and smartphones and to envision the experience of using a smartwatch. Cecchinato et al. [9] interviewed a small number of early adopters who own smartwatches and investigated how they had been using them in real life. Pizza et al. [42] distributed a smartwatch to each of 12 participants and asked them to use it for 34 days. In the final three days, the participants were asked to capture their usage behaviors with wearable cameras, and the captured videos were analyzed to characterize usage behaviors. Lundell and Bates [31] surveyed 90 Apple Watch users to understand user experience journeys over the four month period.

These studies revealed that most smartwatch usage occurs when users want to check notifications [48] and the current time and to track daily activities [42]. The participants reported that notification checking is very useful, particularly in situations in which smartphone usage is socially unacceptable [9, 42, 48]. When compared with smartphones, glancing at a watch is considered a less obtrusive and disruptive way of checking incoming information while conversing or performing tasks at hand [42]. People generally preferred to interact with smartphones and smartwatches in close proximity rather than larger screens for handling notifications [53]. In some cases, smartwatches were used as extra/second screens (e.g., typing a text message code) [42]. Nonetheless, users often envisioned the experience of using smartwatch as only wearing a watch without carrying their smartphone [32], while in reality they always had to carry their phone for the smartwatch to be *usable* due to wireless connectivity constraints and the limited interaction capabilities of smartwatches [48]. Due to such dependencies, most participants used similar applications on their watch as on their smartphone [48], and smartwatch usage was positively affected by the usefulness of their smartphone [15]. Wearing habits of smartwatches could be similar to those of conventional digital watches [32]: never removing (20%), taking off while sleeping/bathing (47%), and not wearing at home (35%). A longitudinal user experience journey study showed that 6.6% of users stopped using the Apple Watch after four months, and there are three types of smartwatch users: communicators for enhancing communications, tool techies for exploring various smartwatch features (e.g., fitness tracking, payment), and detractors for failing to find compelling values of smartwatches [31].

Min et al. [36] investigated the battery concerns of smartwatches. Users generally treated smartwatches as an auxiliary device of smartphones. Inconvenience due to running out of smartphone battery was greater than that of smartwatch battery (30% vs. 78%). Interestingly, many users kept wearing their watches even though batteries ran out. Users frequently interacted with their smartwatches. Overall, interaction patterns are quite different from conventional smarpthone usage [27]. In their three week measurement of 19 Android Wear users (mostly G-Watch), they found that mean interaction duration and frequency was 20 minutes and 96 times per day, respectively; and mean session duration was 13 seconds (38% of sessions lasted less than 5 seconds). Users mainly recharged smartwatches before sleeping with mean inter-charge duration of 31 hours.

Besides functional aspects, users also considered aesthetic of smartwatches, such as the color and shape of the watch, to be important for purchasing and wearing decisions [32, 48]. Users wanted their smartwatches to look like an ordinary wristwatch, not like a noticeable gadget on their wrist, and they preferred a design that suits both casual and formal situations [32]. A recent user preference study affirms this statement in that display shape is more critical than brand and price [19]. In reality, it is somewhat difficult to find a smartwatch that blends well with a formal or stylish outfit, and as a result, they chose an inconspicuous design with a subdued color, instead [48]. According to recent psychology studies, researchers found that smartwatch usage is related to self-expressiveness, a concept referring to the use of some products critical to their social identity [10], and aesthetic importance may be attributed to individual differences in their perception of smartwatches as technology or fashion accessories [11].

### 2.2 Activity Tracking Behaviors

Fitness tracking is one of the key features of a smartwatch. While prior studies mostly focused on user experiences of fitness trackers, such as the FitBit, not much research has been directed at understanding how people use health and fitness tracking functions in a smartwatch [39]. In this section, we review some representative studies about activity trackers to provide background knowledge of how a smartwatch can be used for health and fitness management. The literature identified that a major reason for using wearable trackers is to quantify daily activities. Personal informatics encompasses this kind of personal tracking and studies how people interact with devices and data and for what purposes. Li et al. [28] proposed a stage-based model for personal informatics that consists of five different stages, namely preparation, collection, integration, reflection and action, and uncovered various barriers in personal informatics (e.g., types of data and interoperability), which often cascade into different stages. Rooksby et al. [47] presented a perspective on activity tracking as *lived informatics*, in which people find meaning in their daily lives through the information they get from activity tracking. Another work of Li et al. [29] presented the typologies of questions that people often ask about their data for reflection (i.e., current status, history, goals, discrepancies, surrounding context, and other factors). The questions users value change depending on what phase of reflection they are situated in (e.g., discovery and maintenance), and people move back and forth between discovery and maintenance phases.

While these studies investigated about the behaviors of people while they are using personal trackers, some recent studies focused on the process of people abandoning their activity trackers, and their lives after abandonment. Clawson et al. [12] studied the reasons why people abandon their health tracking devices by analyzing the advertisements uploaded to Craigslist by people to sell their tracking devices. Some people abandoned their device because they accomplished their goals, and some because of the change in activity and health status. However, more than a quarter of users were abandoning their devices because their expectations and actual usage experience did not match. Epstein et al. [13] extended prior studies by studying not only about why people abandon their devices, but also how their lives change after the abandonment. There are several additional reasons of stopping tracking such as the cost of data collection and management, discomfort with information and data accuracy concerns. After abandoning their devices, some people were indifferent, but some felt guilt and frustration with the fact they failed to accomplish their tracking goals. Also, there were some positive changes after abandonment, such as feeling freedom as they are out of bothersome trackers, or continue to use knowledge they acquired from the experience of tracking. Motti and Caine [37] analyzed user reviews of wearable devices (both watches and bands) in Amazon and found that most interaction problems are attributed to the platform issues (e.g., tracking accuracy, usability issues, synchronization, and battery), which contribute to frustration and interruption and may result in abandonment.

Prior studies also investigated actual usage and abandonment of activity trackers over time, but conflicting results were observed. It is commonly believed that abandonment rate is high. In their six-week study of 26 users, Shih et al. [49] found that 75% abandoned Fitbit trackers after four weeks. In their 10-month in-the-wild trial of activity tracking app called Habito, Gouveia et al. [17] found that 64% of users abandoned the app less than a week. This kind of abandonment was similarly observed in general smart devices as shown by Lazar et al. [26]; 80% abandoned in two months. In contrast, Meyer et al. [34, 35] analyzed two longitudinal datasets (i.e., a one-year-long data from 43 patients, and an in-the-wild dataset from 39 regular users). Unlike prior studies, users consistently wore the trackers until they stopped. In a large-scale study of tracking college students with Fitbits, Purta et al. [43] found that participants were highly compliant with longitudinal data collection with average compliance level of 67%. These conflicting results could be attributed to diversity of users such as readiness of behavioral changes, purpose of usage, and incentive for participation: for example, higher adoption were partly observed among those who were ready to change their behaviors [17, 49], used trackers for the purpose of improving health and exercise [22], and were paid for their participation [43]. In their data analysis, Epstein et al. [14] showed that lapsing (temporary stopping) may lead to multiple short/long trials of trackers. Similarly, Meyer et al. [34, 35] identified various long-term use groups based on usage density such as power users (high wearing density) and intermittent users (low wearing density) and usage trial frequency such as experimenter, lapsing users, and consistent users. In addition, most users fall in one or two of usage patterns and this pattern did not change over time.

## 3 MEASUREMENT ANALYSIS OF SMARTWATCH WEARING BEHAVIORS

Analyzing how people wear their watches over time is the first step towards understanding smartwatch wearing behaviors. One of the best ways to study this is to analyze actual sensor data from smartwatches and to characterize detailed temporal wearing behaviors. Towards this goal, we distributed Apple Watches to 50 participants and collected the activity tracker data of the Apple Watch (i.e., heart rate and step count) for 203 days. In this section, we propose a machine learning approach to determine a user's smartwatch wearing status with the activity tracker data. We then present the quantitative results of smartwatch wearing behaviors, such as daily wearing patterns, diurnal wearing patterns, and take-off patterns.

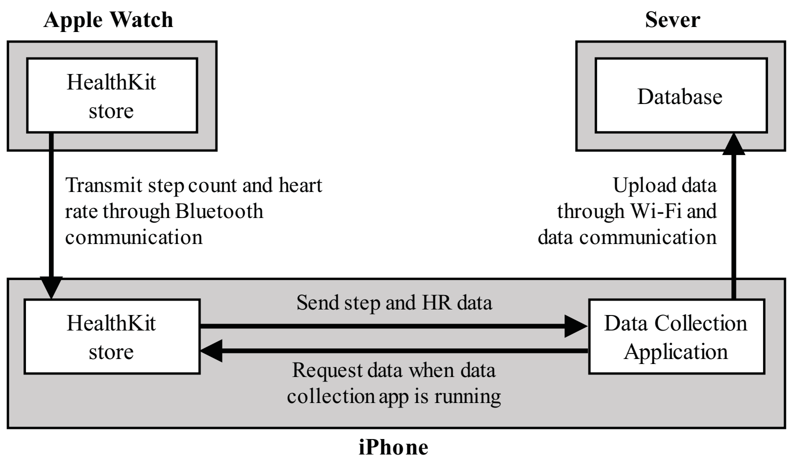


Fig. 1. An overview of the data collection system

## 3.1 Data Collection Method

*3.1.1 Apparatus.* We collected the dataset for behavior analysis from Apple Watch users. We developed a custom data collection system as shown in Figure 1. When a user is wearing her Apple Watch, it periodically collects her step count and heart rate through its embedded sensors. The collected data is automatically stored in an encrypted database called HealthKit store. When the watch is connected with the user's iPhone through Bluetooth, the sensor data in the HealthKit store of the Apple Watch is transmitted to the iPhone's HealthKit store. We developed an app that periodically retrieves sensor data from the iPhone's HealthKit store and transmits the data to a remote server. Our app runs in the background and the fetch interval was set based on the iOS API with the following option: *UIApplicationBackgroundFetchIntervalMinimum*, which is known to execute every 10 minutes or longer [2].

*3.1.2 Participants.* We recruited 50 participants who were willing to use the Apple Watch for a long period of time. Recruitment was advertised through online bulletin boards of university community websites. All participants were iPhone users affiliated with a university in South Korea; 36 participants were male, and 14 were female. The age of the participants ranged from 18 to 37 years, with average 25.58 years old (SD=4.29). Eleven participants were undergraduate students and 37 were graduate students. The remaining two participants were a faculty member and staff member of a school.

We provided an Apple Watch to each participant and let them use it as their own device. We instructed the participants to install our custom data collection app on their phone, which automatically uploaded the sensor data (i.e., heart rate and step counts) from their iPhone's HealthKit service to our server. We collected 203 days of usage data from March 28, 2016 to October 16, 2016. The date each participant started to use their Apple Watch varied within the range of one month, and the first participant began use on March 28, 2016. The participants were not compensated with any monetary rewards, but they were allowed to keep their Apple Watch.

We first divided the days into those days during which participants wore their smartwatches and those days during which they did not wear their smartwatches. During the 203 days of data collection, participants wore their smartwatches for 171.88 days on average (SD=41.16). Figure 2 shows the histogram of the number of wearing days. We note that there were four participants who wore their smartwatches for significantly fewer days than the other users. They wore their smartwatches for less than 100 days out of the total 203 wearing days.

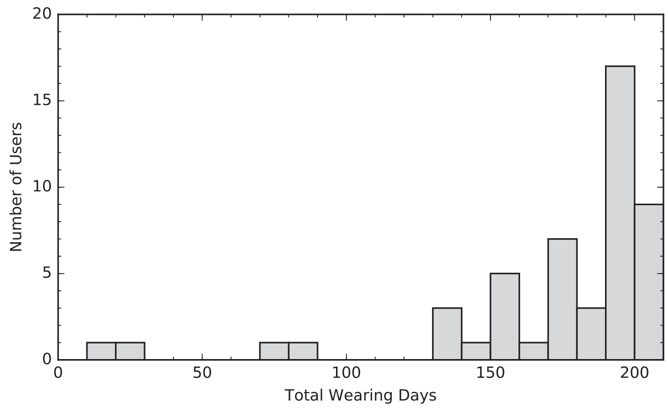


Fig. 2. Distribution of total wearing days of all 50 participants. Four out of the 50 users wear their smartwatches for less than half of the total wearing days and stopped wearing the watch during the first half of the wearing period.

Unlike the other participants, they stopped wearing their watches. Nonetheless, we found that their wearing behaviors are similar to the other participants, and thus, we did not remove their data in our analyses.

## 3.2 Smartwatch Wearing State Recognition

We present a smartwatch wearing state recognition model that aims to infer whether a user is wearing a smartwatch. Recognizing this solely based on the activity tracker data, namely step counts and heart rates from the Apple Watch is non-trivial for at least the following reasons: (1) in the Apple Watch, heart rates may not be available when there are significant movements (e.g., walking and exercising); (2) step counts are collected even though a user is not wearing her watch; and (3) having wireless connectivity between smartphone and watch does not mean that a user is currently wearing her watch. For wearing state recognition, we collected the ground truth dataset to analyze the characteristics of data collection. From this, we trained and evaluated various machine learning models for accurate recognition. Furthermore, we collected a week-long dataset to test whether our model scales to everyday scenarios.

*3.2.1 Ground Truth Data Collection.* For state recognition, we need a ground-truth dataset of smartwatch wearing behaviors, because we do not know the ground truth states of the data collected through our collection system. We used a scenario-based ground truth collection method that is widely used in activity recognition literature [45]. In the scenario-based method, participants simply performed a set of pre-scripted activities for data collection.

We hired six participants for the ground truth data collection. Among the six participants, three participants were male, five participants were in their 20s, and one participant was in the 30s (Mean=26.17, SD=2.48). We recruited them from our campus, and they were not participants in our longitudinal data collection experiment. According to our scenario, a participant was asked to conduct a series of seven activities, which take about 2 hours. We compensated each participant with 30,000 KRW, which is approximately 25.0 USD.

We carefully chose representative activities on the university campus (see Table 1); the labels of these activities include Take-off, Charge, Study/Work, Walk, Eat, Rest, Sleep, and Exercise. There were three activities for which participants did not wear their watches: Take-off, Charge, and one instance of Walk. We had two different walking activities: one wearing and the other without wearing the smartwatch. For the Take-off and Charge activities, we

Table 1. Activity process and details of ground truth data collection experiment

| Activity | Wearing State | Activity Details | Location | Time |
|---|---|---|---|---|
| Take-off | Not-wearing | Put Apple Watch still on a desk | Experiment place (Lounge) | 25 min. |
| Charge | Not-wearing | Plug Apple Watch into a charger | Experiment place | 25 min. |
| Study | Wearing | Study or work in front of a desk | Experiment place | 25 min. |
| Walk | Wearing | Take a walk to school restaurant | Inside campus | 25 min. |
| Eat | Wearing | Have a meal in the school restaurant | School restaurant | Adjust to eating time |
| Walk | Not-wearing | Take a walk back to the experiment place | Inside campus | 25 min. |
| Rest | Wearing | Take a rest as participants usually do | Experiment place | 25 min. |
| Exercise | Wearing | Exercise by following the provided workout video | Experiment place | 25 min. |

collected the data before the participant arrived at the experiment site, and we asked the participants to perform the rest of the activities in the table in sequence. Regarding the Exercise activity, we asked participants to watch a video of a full-body workout that did not require any equipment and to follow the instructors; examples of the exercises include jumping jacks, pushups, and crunches. In this case, we instructed the half of the participants to enable the Workout App. The data collection period of each activity was set to 25 minutes, except for Eat. We allowed the participants to take their time when eating their meals.

*3.2.2 Data Characteristics.* We analyzed the ground truth data to understand the basic characteristics of the activity tracker dataset, by examining (1) whether step counts are collected under not wearing conditions; and (2) whether heart rates are properly collected under various activities.

First of all, our results showed that the Apple Watch collected step counts even though a user does not wear the device. In Figure 3, we present how step counts are collected over time under different conditions: (1) high activity with wearing (walking), (2) low activity with wearing (eating), and (3) high activity without wearing (walking with a watch in the pocket). When a user wears their smartwatch during a state of high activity, it collects the step count *every minute*. However, when a user is in a low-level activity state (e.g., eating or staying in the office), it collects the step count at irregular intervals. The average step count collection interval was 8.49 minutes (SD=2.30). In this example, the user was having a meal in a restaurant during this low-level activity period. Step counts were also collected in situations in which a user was not wearing their watch but moving, i.e., returning from a cafeteria with their watch in the pocket.

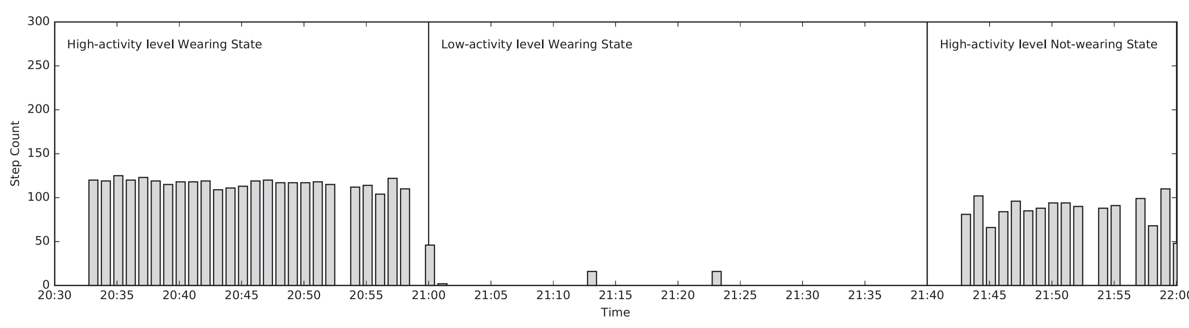


Fig. 3. Example of step count data over time

Apple's HealthKit module treats the same data types measured from multiple trackers equally. When multiple trackers are used, there may be duplicate measurement values stored in the HealthKit. In our case, step counts were collected from both the smartphones and smartwatches. Suppose that a user takes a walk for 30 minutes; then step counts measured by the Apple Watch and iPhone will be stored in the HealthKit data storage. When retrieving the data from the HealthKit storage, we have a list of data items, each of which has a time stamp and measured step count without source labels, and thus, there will be duplicate step counts. By analyzing the dataset, we found that the data sampling frequencies of the smartwatch and smartphone differ and that this information could be used to filter duplicate measurements. Our data analysis revealed that step counts from the Apple Watch had relatively short sampling intervals with an average value of 2.03 minutes (SD=2.67), whereas the step counts from the iPhone had longer sampling intervals with an average of 8.65 minutes (SD =2.28). Due to the time interval difference, the magnitude of step counts differs depending on the data source. For example, the step counts from the Apple Watch are 113.03 on average (SD=156.53), and those from the iPhone are 567.22 on average (SD=400.03). Thus, it is possible to filter out measurements from the iPhone with a simple threshold method; that is, we remove step counts with values greater than a threshold value. For a given training dataset, we can easily find an optimal threshold value that maximizes the decision accuracy by using Decision Tree methods, in which an optimal decision boundary can be found by iteratively searching for a threshold value that most reduces the unorderedness (or entropy) or that has the highest information gain. Using our training dataset, we found that information gain was maximized when the threshold value was set to 207.5. In our follow-up data analysis, we used this threshold to eliminate duplicate step counts that were generated by iPhones.

We then examined how heart rates were collected under various situations. A tuple of data (timestamp and heart rate) was periodically collected from the HealthKit storage. According to the specification [1], the Apple Watch measures a user's heart rate when the user becomes stationary without moving much. According to our dataset, heart rates were sampled every 5.62 minutes on average (SD=3.05) when users' activity level was low.

In Figure 4, we plot the heart rates of a participant from 16:30 to 17:25 during the day of experimentation. According to our ground truth data, the participant was wearing the Apple Watch and worked on a laptop during the period of low activity (till 17:00). Observe that the heart rate collection intervals are irregular, and that step counts are not collected in this period (due to low activity levels). In contrast, when a user became active with significant bodily movements (say due to walking), we found that heart rates were not collected. For example, in Figure 4, there were no heart rate samples when the user started walking (as of 17:00), but step counts were updated almost every minute.

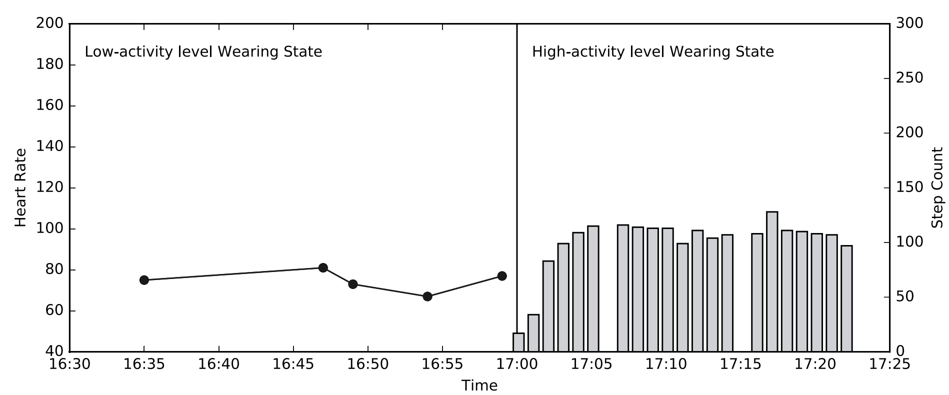


Fig. 4. Heart rate and step count data over time

There was an exception for heart rate data collection. The Apple Watch has an exercise tracking app called Workout, which forces the watch to collect heart rates fairly regularly [1]. When this app was running, we found that the sample data were collected almost every minute. Figure 5 contrasts data collected with and without the Workout app enabled, while the user was performing the same activity—we asked the user to watch an exercise video and mimic the activity. As shown in Figure 5, the heart rate was regularly sampled only when the Workout app was enabled.

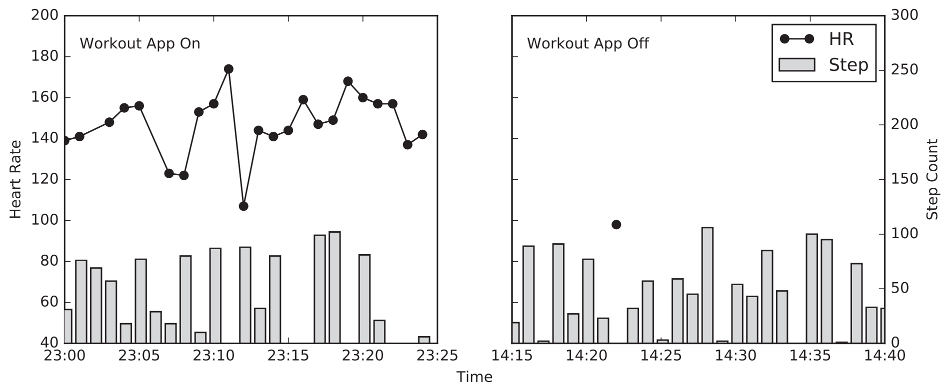


Fig. 5. Heart rate and step count data over time with conditions of Workout application on (left) and off (right)

*3.2.3 Wearing State Recognition Model and Evaluation Results.* We first segment the time series dataset from the Apple Watch (heart rates and step counts) with overlapping sliding windows. For a given window of $t$ minutes, we extracted six features, namely *step count mean, step count variance, heart rate mean, heart rate variance, number of step count samples*, and *number of heart rate samples*. The mean and variance feature types were selected because these features were typically used in existing activity recognition models [7]. Furthermore, collection intervals of the step counts and heart rates from the Apple Watch is irregular, and thus, we expect that the number of samples per window (i.e., the number of step count samples and number of heart rate samples) varies widely depending on whether a user is wearing the watch.

For each feature vector extracted from a window of data, we use an instant classifier to judge whether a window belongs to a wearing or not-wearing state. There may be errors in classification; for example, a user may have continued wearing their watch, but there were few errors in the middle, and thus, the classifier might incorrectly judge that there were several take-off events. For robust classification, we adopted a Discrete Hidden Markov Model (DHMM), which has often been used to overcome erroneous classification results in human activity recognition [30, 45]. A DHMM uses the transition probability from one state to another as one of its parameters for classification; in this way, sudden errors can be easily corrected because such transitions are less likely to occur according to the probability transition matrix.

We consider various instant classifier algorithms that are frequently used in activity recognition: Decision Trees (DT), Support Vector Machines (SVM), Nearest Neighbors (NN), and Random Forests (RF). Performance evaluation was conducted with the leave-one-user-out cross validation method. Since performance depends on the window size, we considered various window sizes, ranging from 2 to 30 minutes in intervals of 2 minutes and selected an optimal window size that maximizes precision and recall metrics. Precision in the recognition of the wearing state was calculated using the ratio of numbers of true wearing states from the total number of wearing states judged by an instant classifier. The recall was calculated using the ratio of numbers of wearing states that were correctly classified from numbers of actual wearing states. In our model, we first selected the best performing instant classifier and then applied a DHMM to the selected classifier.

Table 2. Precision with different classifiers and window sizes

| Window size (min.) | DT | SVM | NN | **RF** | **RF-DHMM** |
|---|---|---|---|---|---|
| 2 | 0.839 | 0.833 | 0.767 | 0.850 | 0.851 |
| 4 | 0.862 | 0.838 | 0.872 | 0.876 | 0.880 |
| 6 | 0.866 | 0.814 | 0.859 | 0.879 | 0.883 |
| 8 | 0.907 | 0.812 | 0.843 | 0.888 | 0.900 |
| 10 | 0.854 | 0.784 | 0.872 | 0.861 | 0.832 |
| 12 | 0.853 | 0.774 | 0.882 | 0.863 | 0.854 |
| 14 | 0.878 | 0.747 | 0.865 | 0.888 | 0.891 |
| **16** | 0.915 | 0.743 | 0.871 | **0.931** | **0.924** |
| 18 | 0.873 | 0.745 | 0.844 | 0.898 | 0.864 |
| 20 | 0.908 | 0.729 | 0.846 | 0.908 | 0.881 |

Table 3. Recall with different classifiers and window sizes

| Window size (min.) | DT | SVM | NN | **RF** | **RF-DHMM** |
|---|---|---|---|---|---|
| 2 | 0.522 | 0.598 | 0.621 | 0.535 | 0.808 |
| 4 | 0.646 | 0.748 | 0.657 | 0.646 | 0.857 |
| 6 | 0.749 | 0.846 | 0.747 | 0.763 | 0.827 |
| 8 | 0.795 | 0.892 | 0.822 | 0.788 | 0.898 |
| 10 | 0.791 | 0.932 | 0.773 | 0.861 | 0.944 |
| 12 | 0.778 | 0.977 | 0.866 | 0.879 | 0.900 |
| 14 | 0.784 | 0.965 | 0.779 | 0.855 | 0.932 |
| **16** | 0.876 | 0.984 | 0.820 | **0.923** | **0.939** |
| 18 | 0.765 | 0.991 | 0.845 | 0.897 | 0.924 |
| 20 | 0.822 | 0.990 | 0.744 | 0.881 | 0.939 |

Table 4. Accuracy of RF-DHMM

| | |
|---|---|
| User 1 | 0.923 |
| User 2 | 0.929 |
| User 3 | 0.931 |
| User 4 | 0.885 |
| User 5 | 0.840 |
| User 6 | 0.963 |
| **Average** | **0.912** |

For the different classifiers and window sizes, we present the precision results in Table 2, and the recall results in Table 3. According to our results, the RF with a 16-minute window size yielded the best result, with 93.1%

precision and 92.3% recall rates. The SVM showed better recall results, but it had very low precision. The use of a DHMM resulted in slight increment of recall at the cost of slight decrement of precision. Thus, we chose the RF for classification and set the window size of 16 minutes for the rest of the analyses, and the classification accuracy is presented in Table 4. Here, the window size of 16 minutes means that the model can classify the states in every 8 minutes due to overlapping slide windowing.

*3.2.4 Model Validation with Everyday Scenarios.* So for we built and tested our model based on the dataset that we collected via a scenario-based ground truth collection method where participants simply performed a set of pre-defined activities (e.g., study, walk, eat, rest, sleep, exercise). For generalizability of the proposed model, it is necessary to test the model by collecting a naturalistic wearing data. However, it is quite challenging to collect such a dataset because ground-truth logging is cumbersome, and the dataset is heavily skewed to the wearing state (only a few take-off events per day). To address this challenge, we decided to use (1) an activity logging app (Toggl) to track users' daily activities as well as their wearing states, and (2) an experience sampling app (Paco) to periodically generate take-off events. We collected a week-long dataset from four users. Our model evaluation with this dataset showed that even in this everyday scenario, our model is highly accurate, achieving average classification accuracy of 0.972.

*Data collection method* Our goal is to collect users' activities and their wearing states in their everyday lives. For activity logging, Toggl was used. Whenever an activity transition happened, the participants were asked to label the current activity, mark the current wearing status (i.e., wearing, not wearing, charging), and start a timer. After finishing the activity, the user cancelled the timer. This procedure was repeated throughout the day while they were awake (i.e., excluding sleeping hours). In addition, Paco App was used to generate a series of take-off and wearing events. This is because in a naturalistic setting, we have too few instances of take-off events. For a given day, we had a fixed schedule of 5 take-off events (every 4 hours as of 10:20AM), where each take-off event is followed by a wear event after 30 minutes (i.e., 10 events per day). Upon receiving these alarms, we asked the users to log them using Toggl and to react to the request accordingly. To avoid situations in which a take-off event happens across multiple activities, we asked the participants to accept a take-off request from Paco App only if their current activity can last longer than 30 minutes. Besides these scheduled take-off events, there could be other voluntary take-off events (e.g., charging their watches or taking a shower). We hired four participants in their 20s (mean: 23.0 SD: 3.5; 1 female), which were different from our first data collection study. The data collection lasted for a week. We compensated each participant with 80,000 KRW, which is approximately 66.7 USD.

*Results* All the participants finished data collection for a week. The average number of daily take-off events accepted per user was 4.14 (SD = 0.77), and it was consistent over time. The average hours of wearing and take-off states per participant were given 8.84 (SD = 2.22) and 3.18 (SD = 1.48), respectively.

For each user, we performed an affinity diagramming on the activity labels. We found that our participants performed diverse activities. Our participants spent most of their time on typical daily activities such as study, computer work, eating, walking, socializing, resting, and class attendance, but there were also a range of other activities such as TV watching, cleaning, bike/vehicle riding, and shopping. For example, P2 spent about 18 hours for study and 8 hours for walking. During these activities, there were 8 and 4 take-off events, respectively. The average wearing and not wearing duration of the study activity were given as 65 minutes (SD = 44) and 41 minutes (SD = 11).

Since the duration of the wearing state is longer than that of the not-wearing state, we used random sampling that modifies the imbalanced dataset to provide a balanced distribution. According to a review article for learning from imbalanced data [18], training classifiers with a balanced dataset is known to improve overall accuracy as opposed to that with an imbalanced dataset. In our study, we used a random oversampling (or up-sampling) method on the take-off state data in order to remedy imbalance. We applied our trained model (based on our scenario-based dataset) to this balanced dataset to check how accurately it can identify wearing states. Note that there is a possibility of overfitting in random oversampling, because replicas are appended into the existing dataset [18]. Exploring various kinds of sampling techniques for imbalanced learning is beyond the scope of this work, and we leave that as part of future work.

Table 5. Classification results

| | Precision | Recall | Accuracy |
|---|---|---|---|
| P1 | 0.950 | 0.966 | 0.958 |
| P2 | 0.965 | 0.952 | 0.959 |
| P3 | 0.987 | 0.972 | 0.979 |
| P4 | 0.986 | 0.975 | 0.980 |
| Total | 0.978 | 0.967 | 0.972 |

Table 5 shows that our algorithm can accurately classify the wearing states even in everyday scenarios as its accuracy was over 0.95. We examined what were the typical patterns of the classification errors. We found that most of the false negatives were related to erroneous heart rate sensing with the Apple Watch. For example, P1 was studying while wearing the watch, but the Apple watch failed to sense P1's heart rates. We found that false positives were mostly related to false heart rate reporting from the Apple Watch and HMM's correction errors. For example, while socializing, P2 took off the watch and put it on the pocket, but there were heart rate data reported, which resulted in false classification.

## 3.3 Wearing State Analysis Results

We analyzed our activity tracker dataset and performed wearing state recognition by following the procedure in Section 3.2. In the following, we provide our analysis results of smartwatch wearing behaviors; e.g., daily wearing patterns, diurnal wearing patterns, and wearing/taking-off event occurrences in a day.

*3.3.1 Daily Wearing Patterns.* The results show that our participants wear their watches for 10.48 hours a day on average (SD=3.47). As shown in Figure 6, wearing times on weekdays and weekends (including holidays) have different tendencies. The average wearing times are 11.32 hours (SD=3.53) and 8.66 hours (SD=3.60) on the weekdays and weekends, respectively. It is interesting to note that our participants tended to wear their smartwatches longer on weekdays than on weekends.

*3.3.2 Diurnal Wearing Patterns.* We examined what time of the day people wear their watches, which gives us a wealth of information about wearing habits. For each participant, we calculated a 24-dimension probability vector, where each attribute was calculated over the entire period as the fraction of time during which a user wore a smartwatch in that hour. An example vector in Table 6 shows that the user had a 78% chance of wearing their watch between 0 AM and 1 AM and a 59% chance between 1 AM and 2 AM.

Table 6. Example of a smartwatch wearing probability vector. Each attribute value represents the probability of wearing the smartwatch during the designated hour

| Attribute Number (Time) | 0 | 1 | 2 | ... | 22 | 23 |
|---|---|---|---|---|---|---|
| Wearing Probability Vector | (0.78, | 0.59, | 0.68, | ... | 0.79, | 0.64) |

This vector notation allows us to cluster similar users into groups. We used the spectral clustering algorithm, which is known for providing effective clustering of high dimensional data with non-convex boundaries [51]. For

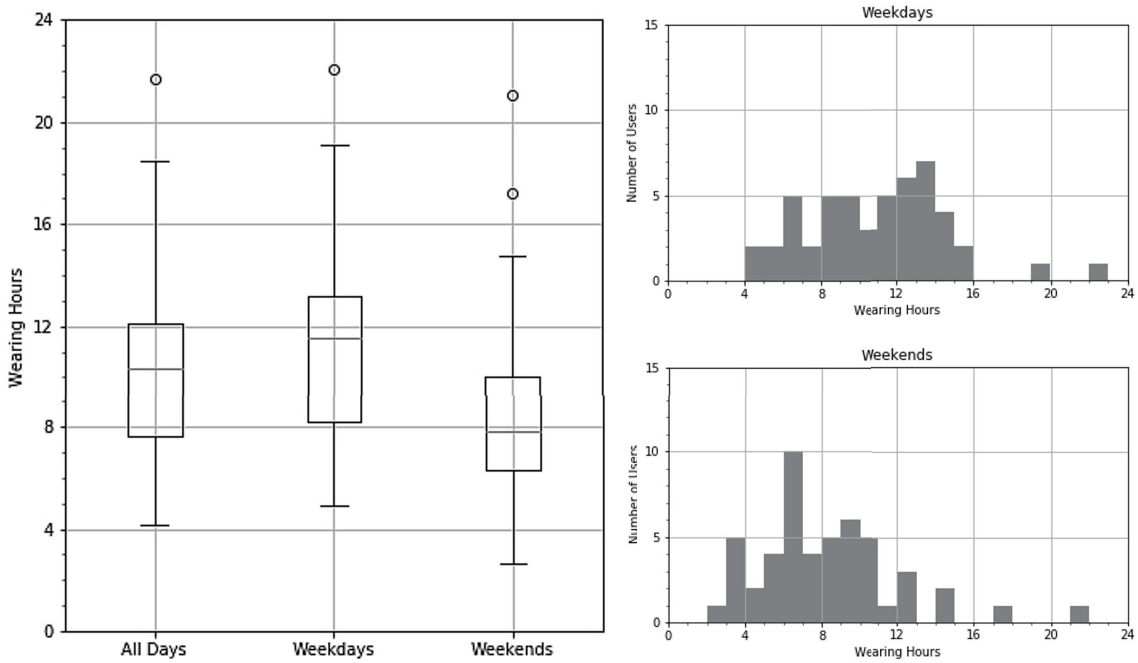


Fig. 6. Smartwatch wearing time distribution of all days, weekdays, and weekends (including holidays)

spectral clustering, we first built a cosine similarity matrix using each user's probability vector. We then applied the spectral clustering algorithm by varying the number of clusters. We found a reasonable grouping when the number of clusters was set to three.

| Time | 0-1 | 1-2 | 2-3 | 3-4 | 4-5 | 5-6 | 6-7 | 7-8 | 8-9 | 9-10 | 10-11 | 11-12 | 12-13 | 13-14 | 14-15 | 15-16 | 16-17 | 17-18 | 18-19 | 19-20 | 20-21 | 21-22 | 22-23 | 23-0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Work-hour wearer | (0.09, | 0.05, | 0.04, | 0.03, | 0.03, | 0.03, | 0.03, | 0.06, | 0.14, | 0.30, | 0.42, | 0.50, | 0.54, | 0.53, | 0.52, | 0.52, | 0.51, | 0.52, | 0.50, | 0.46, | 0.41, | 0.34, | 0.27, | 0.16) |
| Active-hour wearer | (0.28, | 0.18, | 0.11, | 0.07, | 0.05, | 0.04, | 0.05, | 0.05, | 0.10, | 0.23, | 0.36, | 0.45, | 0.50, | 0.53, | 0.55, | 0.56, | 0.56, | 0.57, | 0.55, | 0.54, | 0.52, | 0.49, | 0.44, | 0.38) |
| All-day wearer | (0.45, | 0.45, | 0.39, | 0.35, | 0.33, | 0.30, | 0.30, | 0.30, | 0.31, | 0.34, | 0.37, | 0.40, | 0.41, | 0.43, | 0.48, | 0.48, | 0.48, | 0.49, | 0.50, | 0.51, | 0.51, | 0.48, | 0.49, | 0.46) |

Table 7. Representative smartwatch wearing probability vector of three user groups

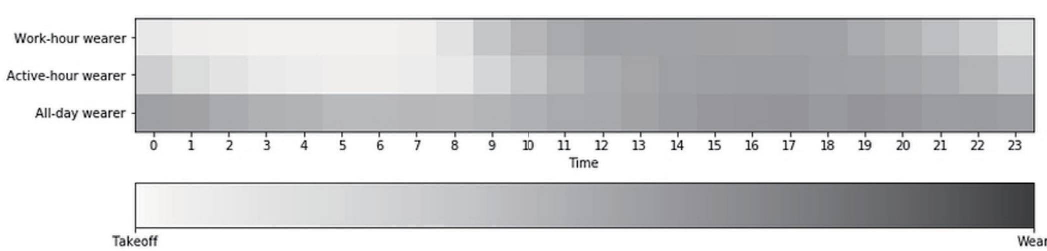


Fig. 7. Visualization of three different user groups based on smartwatch wearing times

The representative wearing probability vector for each group is shown in Table 7, and its heat map is presented in Figure 7. The representative vector value was calculated by averaging the vectors of all participants inside that group. We named the three user groups as work-hour wearers, active-hour wearers, and all-day wearers.

The work-hour wearer group includes the type of users who tend to wear their smartwatches while they work (n=29). For example, during work hours (from 9 AM to 6 PM), on average they had a 48.44% chance of wearing it (SD=7.75). Unlike other groups, their wearing chances considerably drop after work hours: 6-9PM: 45.67% (SD=4.51), and 9PM-12AM: 25.67% (SD=9.07) (vs. during work hours: 9AM-6PM: 48.44%, SD: 7.75). One possible explanation is that the users take off their watches when they return to their home.

The active-hour wearer group (n=15) is quite similar to the work-hour wearer group, but they wear their watches as long as they are active. Unlike the work-hour wearer group, their wearing chances did not change significantly even after work hours: 6-9PM: 53.67% (SD=1.53), and 9PM-12AM: 43.67% (SD=5.51) (vs. during work hours: 9AM-6PM: 47.89%, SD: 11.54). In both the work- and active-hour groups, the chances of wearing considerably drops after midnight, which means that most users take off their watches before going to bed.

The last group is the all-day wearer group (n=6) whose wearing behavior is similar to the active-hour wearer group: all-day wearers (9AM-6PM): 47.89% (SD=7.75) vs. active-hour wearers (6-9PM): 53.67% (SD=1.53). The key difference is that the chances of wearing smartwatches during sleeping hours were higher: all-day wearer (12AM-9AM): 35.33% (SD=9.07) vs. active-hour wearers (12AM-9AM): 10.33% (SD=7.97).

*3.3.3 Take-off Events.* We investigate the take-off periods in the time series dataset of wear/take-off states. When wearing states do not occur for more than 8 minutes, we speculate that a take-off event had occurred. The 8-minute time interval is determined based on the state recognition frequency from the window size. Meanwhile, sleeping hours are treated as exceptional take-offs as they occur periodically every day. As most people tend to be asleep at or after midnight, take-offs that occur in this window are deemed as special cases of taking-off for sleep, and thus are excluded from the analysis. In Figure 8 we plot the distribution of the average number of take-off occurrences per day. Our participants (n=50) took off their smartwatches 3.17 times a day on average (SD=1.11). Over the entire data collection period, the participant (P38) with the fewest take-off actions averaged 0.29 take-off events per day (SD=0.89), and the participant (P14) with the most frequent take-off actions averaged 6.83 times a day (SD=2.80). As shown in the average take-off count distribution in Figure 8, most participants (33 out of 46) took off their smartwatches two to four times a day.

We also plot the distribution of take-off durations for each individual in Figure 9. Take-off durations vary widely, ranging from less than 30 minutes to more than 5 hours. The figure shows that most of the take-off events last for 30 minutes to 1 hour. The average take-off duration for all participants was 55.30 minutes (SD=23.77).

*3.3.4 Temporal Wearing Behaviors.* To analyze longitudinal wearing behaviors, we use the following metrics:

- Average daily wearing hours per day of week: for a given day of week (i.e., Monday to Sunday) and a user, we can calculate the average daily wearing hours.
- Average daily wearing hours per week: for a given week and a user, we calculate the average daily wearing hours.
- Break length: for a given user, break length denotes a number of consecutive days of not wearing a smartwatch (i.e., zero wearing hour).
- Streak length: for a given user, streak length denotes a number of consecutive days of wearing a smartwatch (i.e., non-zero wearing hours).
- Usage density: for a given user, usage density denotes the ratio of the number of wearing days to the entire days tracked; i.e., the number of wearing days / (last day of use − first day of use + 1).

Figure 10 shows a user's wearing behaviors for 9 days. For each day, it shows whether a user wore the watch or not: O denotes wearing and X denotes not wearing. In this figure, we have 1 streak instance with its length of

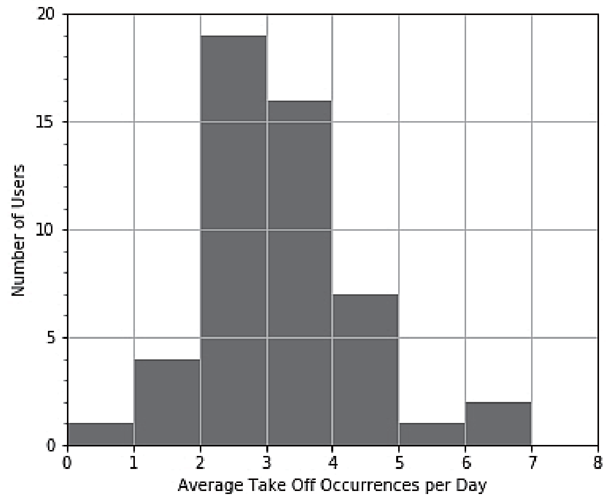


Fig. 8. Histogram of average take-off occurrences per day/user

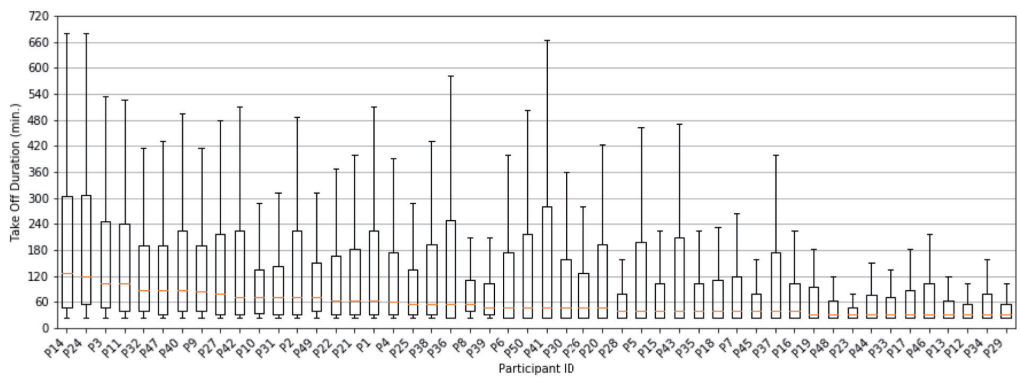


Fig. 9. Boxplots of take-off durations of individual users. Users were sorted based on median take-off duration

5 days (say weekdays) and 1 break instance with its length of 2 days (say weekend). Assuming that the entire tracked duration is 9 days, this user's density is given as 0.78.

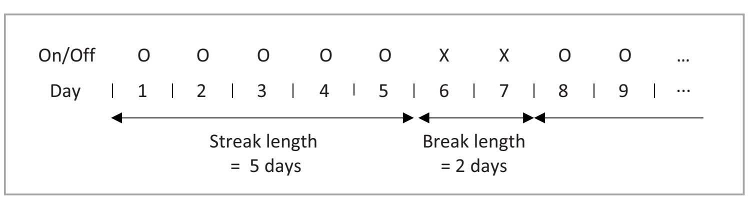


Fig. 10. Illustration of metrics

These metrics allow us to answer the following questions:

- Do wearing hours per day of week show weekly patterns? Do weekly wearing hours significantly decrease over the entire duration?
- How break and streak length distribution look like? Are there any individual variations, and are there any patterns among users?

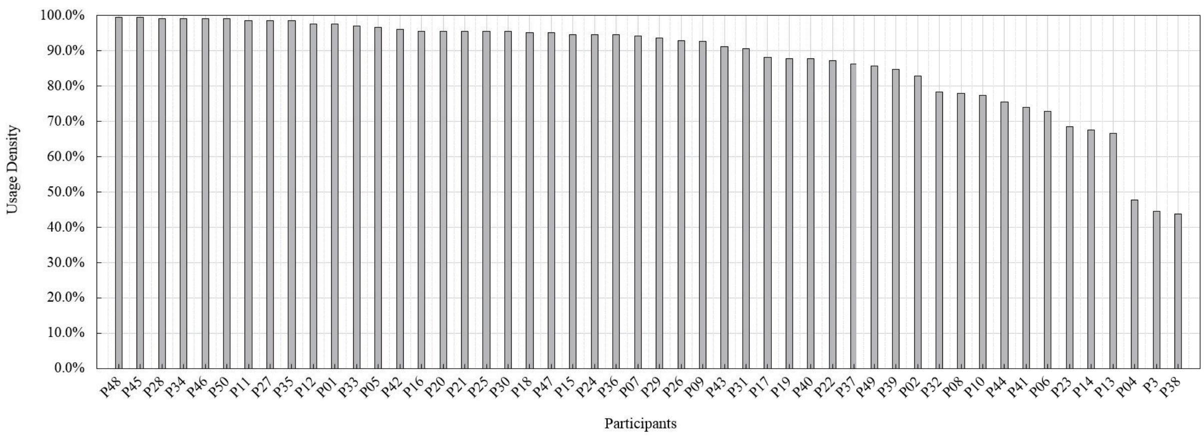


Fig. 11. Usage density per user

Figure 11 shows that usage density of a user is fairly high with its average of 89.8%, which means that on average our participants wore their watches for 6.27 days per week.

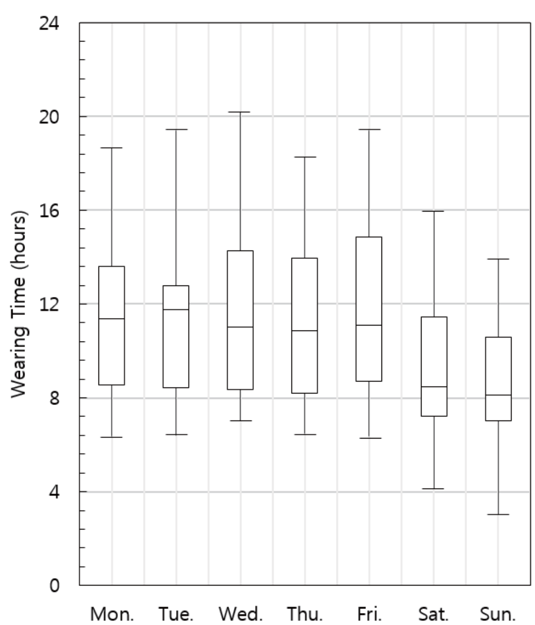


Fig. 12. Average daily wearing hours per day of week/user

Figure 12 shows a weekly rhythm of wearing. The mean wearing hours of Mondays and Tuesdays slightly high with 11.34 (SD: 3.78) and 11.87 (SD: 4.26), respectively. The following days show consistent patterns with their mean values slightly lower than Mondays and Tuesdays. Wearing hours of the weekend are fairly lower than those of the week days.Furthermore, Sundays had slightly lower wearing hours than Saturdays (8. 37 hours with SD 4.57 vs. 8.61 hours with SD 4.45). A paired $t$-test showed that there was a significant difference between the wearing hours of the week days (11.92 SD: 4.25) and those of the weekend days (8.61 SD: 3.92) ($t = 2.489, p = 0.041$). We also examined whether a user's average daily wearing hours per week change over 29 weeks. Our results showed that they were fairly consistent over time; the median wearing hours ranged in between 8.34 and 13.87.

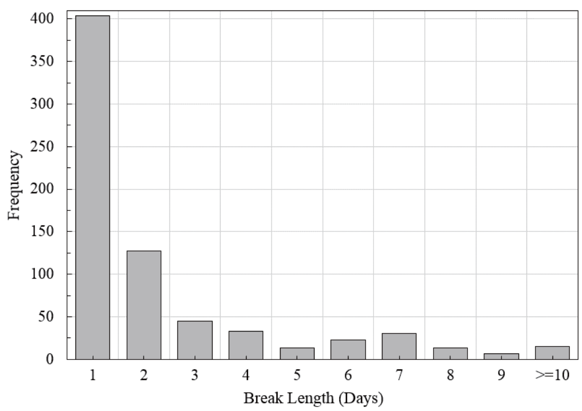


Fig. 13. Break length dist. (Mean: 21.7 days, SD: 15.2)

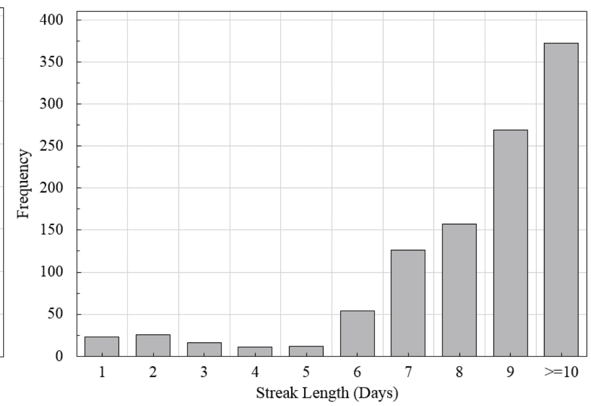


Fig. 14. Streak length dist. (Mean: 25.7 days, SD: 12.3)

Figure 13 and Figure 14 show the break and streak length distribution, respectively. Overall break length distribution follows the geometric distribution, which is a discrete counterpart of the exponential distribution. This means that participants tended to have short break duration of several days. We found significant individual variations of break and streak length distributions.

Figure 15 and Figure 16 show the boxplots of break and streak length per user, respectively. Interestingly, the plots show that there are roughly two groups of users: power users, whose median break duration is just 1 day (n=19, 38%), and casual users, whose median break duration is greater than 1 day (n=31, 62%). The casual users can be further divided into two sub-groups depending on whether median break duration is greater than 5 days: low casualness (n=27, 54%) and high casualness (n=4, 8%). A majority of casual users had low casualness in that they consistently wore their smartwatches over the entire period, but they may not wear their watches everyday and may have longer breaks sometimes.

## 3.4 Discussion: Wearing Behaviors of Abandoned Users

We have four participants who wore their smartwatches for significantly fewer days than the other users. These participants were P38, P26, P04, and P03. Our analysis showed that these participants wore their watches less than 100 days (P38: 85, P26: 78, P04: 20, and P03: 12) because all the participants stopped wearing their watches: P38 = October 2, P26 = June 14, P04 = May 11, and P03 = April 25. Regarding temporal wearing patterns, we found that P26 belongs to the group of power users, P38 and P03 belong to the group of low casualness users, and P04 belongs to the group of high casualness users. For example, P26 had only 6 single-day breaks, but this user stopped wearing the device after two months or so. P38 had three long breaks; and after the third long break, P38 stopped wearing the device. Regarding their diurnal wearing behaviors, P02 belongs to the work-hour wearer

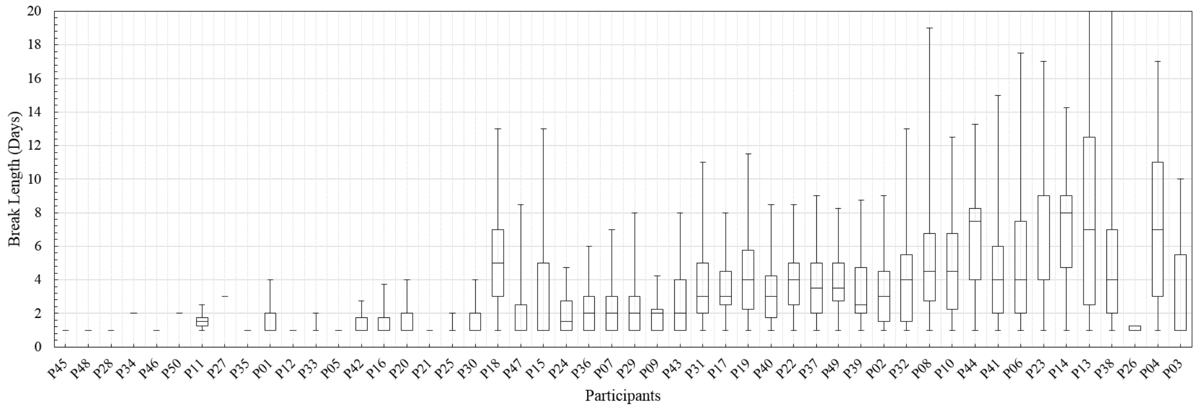


Fig. 15. Boxplot of break lengths per user. Y-axis was truncated at 20. P13 and P38 had the maximum values of 40 and 50 days, respectively.

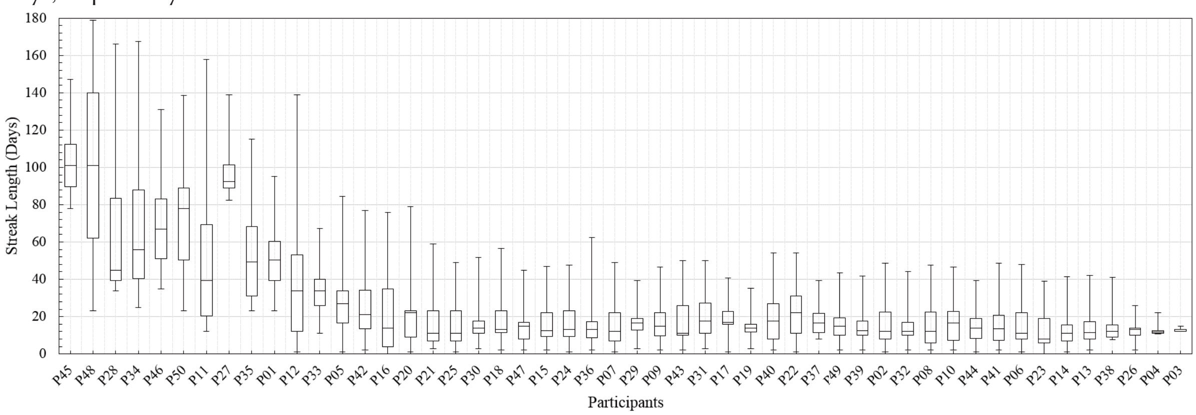


Fig. 16. Boxplot of streak lengths per user. Y-axis was truncated at 180, and P48 had the longest streak length of 179 days.

group, P26 and P04 belong to the active-hour wearer group, and P38 belongs to the all-day wearer group. We did not see any notable patterns of take-off frequency and duration among these participants: P38 and P04 had quite low take-off frequency, whereas P26 and P03 had moderate take-off frequency. Their take-off frequency/duration (minutes) are given as follows: P38 = 0.29 (SD: 0.89) / 99.8 (SD: 159.3), P26 = 3.04 (SD: 2.61) / 191.3 (SD: 268.7), P04 = 0.94 (SD: 1.07) / 118.1 (SD: 173.4), and P03 = 2.79 (SD: 1.68) / 190.4 (SD: 202.8). Our results show that to our surprise, we did not find any notable tendency that wearing behaviors such as daily take-off and streak/break patterns are closely related to abandonment.

## 4 UNDERSTANDING SMARTWATCH WEARING BEHAVIORS

In the previous chapter, we quantified smartwatch wearing behaviors by analyzing the activity tracker dataset collected from Apple Watch users. Through this analysis, we discovered when and how long people do or do not wear their smartwatches. In this section, we aim to corroborate our quantitative findings in Section 3 with the qualitative study. We look at the reasons behind our quantitative findings by investigating a variety of contextual elements, such as places, activities, and social circumstances. Towards this goal, we performed an online survey and in-depth interviews. For temporal wearing behaviors (RQ1), we elaborate on the possible explanations of why such temporal patterns were observed. For factors affecting wearing behaviors (RQ2), we provide major themes uncovered from the content analysis of interview results, such as the need of an immediate response, engagement of in social activities, availability of personal workspace, need for multitasking, smartwatch charging, aesthetics, daily activity tracking, and exercise tracking.

## 4.1 Methods

*4.1.1 Survey Method.* We conducted an online survey with the participants who provided the activity data. Among the 50 participants, 47 participants completed the survey. The survey was conducted nine months after the participants started using the Apple Watches. Therefore, the results and comments obtained from the survey reflect usage experiences and opinions derived from long-term smartwatch use.

After collecting demographic information, we asked about their Apple Watch usage experiences. Information about how often and how long they normally wear their smartwatch was also collected. The survey questionnaire covers four themes: (1) usage purposes and practices; (2) reasons for wearing and not-wearing the watch; (3) smartwatch wearing situations; and (4) take-off events. There were a total of 34 questions. The participants were compensated with 10,000 KRW, which is approximately 8.30 USD.

*4.1.2 Interview Method.* During the survey, we asked the participants whether they would like to participate in a follow-up interview. Twenty people out of the 47 who completed the survey answered that they wanted to take part in the interview. We conducted in-depth interviews with these 20 participants. Of our 20 interviewees, 14 were male. Participants' ages ranged from 20 to 35, and 16 participants were in their 20s. Fourteen interviewees were graduate students, four interviewees were undergraduate students, and the remaining two were a faculty member and a staff member.

The interview was conducted in a semi-structured manner (in Korean). The starting question was always, "Please explain how you typically use your Apple Watch." When participants talked about their daily smartwatch usage, we focused on interesting parts of the story and asked more detailed questions. Commonly asked questions were: "What time do you start to wear and take off your watch?", "How do you use your watch during active hours?", "Were there any times that the smartwatch helped or disrupted your activity?", "Do you exercise regularly? If so, do you use the smartwatch for exercise tracking?", and "What do you think about the design of the Apple Watch?" The survey answers of each participant were used as a reference when conducting the interview. For each person, the interview took approximately 30 minutes, and we compensated participants with 20,000 KRW, which is approximately 16.60 USD.

All responses from the interviews were recorded and transcribed. The scripted answers were grouped into themes with affinity diagramming [6]. The results from the survey were merged into the themes created with the affinity diagram and formed into our findings on smartwatch wearing behaviors. The quotes presented in this paper were translated into English by the authors.

## 4.2 Findings: Understanding Temporal Wearing Behaviors

We provide the key reasons of temporal wearing behaviors through survey and interview data analysis. We present our findings on daily wearing patterns, diurnal wearing patterns, and take-off events.

*4.2.1 Daily Wearing Patterns.* In the survey, we asked the question *"How many hours a day do you normally wear your smartwatch?"* to compare participants' answers with the results from the measurement study. According to the survey, participants wore their smartwatches for an average of 13.38 hours per day (SD=3.64). The measurement analysis results of the survey participants (n=47) showed that they wore their watches for 11.08 hours per day on average, and that the wearing time on weekdays (11.15 hours, SD=3.53) was longer than that on weekends (8.42 hours, SD=3.36). These results show that our participants' subjective reports slightly overestimate their actual wearing hours by about three hours on average ($F = 12.4254, p < 0.001$).

Table 8. Wearing frequency score by place

| Place | Wearing Frequency Score |
|---|---|
| Home, Dormitory | 2.83 (SD=1.47) |
| Workplace, Classroom | 4.57 (SD=0.71) |
| Restaurant, Cafe | 4.51 (SD=0.72) |
| Gym, Sports field | 3.83 (SD=1.36) |

When we asked about wearing hours in the survey, we did not specify whether we wanted to know about wearing hours on weekdays or weekends. Since there are more weekdays, it is likely that our participants estimated their wearing hours mostly based on their experiences on weekdays, for which the wearing hours are slightly longer. In any case, we observe that our participants overestimated their actual wearing hours.

One of the major findings in Section 3.3 was that wearing hours are shorter on the weekends than on the weekdays. Our interview results showed that the participants tended to stay in their homes longer during the weekends, and they were less likely to wear their watches at home. P20 commented, *"There are weekends that I don't go out of my house, and on those days I don't wear the smartwatch. Actually, when stay in my house, I do not use it." (P20)* Most of our participants commented that they usually put on their smartwatches right before they leave their home, as P16 said, *"Since I go out a little later on weekends and I do not wear it at home, I start to wear the watch a little later on weekends" (P16).* Several users commented that wearing a smartwatch is only a routine on weekdays, as P14 said, *"I often forget to wear the smartwatch on weekends and often do not wear it. … After hanging out with my friends, I sometimes totally forget to wear the smartwatch" (P14).* In contrast, many participants responded that they wore their smartwatch if they had specific tasks to do out of their homes, such as personal appointments and grocery shopping: *"I don't use the smartwatch very often on weekends. However, if I have appointments and have to go out, I almost always wear the watch" (P28); "I nearly always don't wear the smartwatch on weekends, but when I go out, like for grocery shopping, I put it on again" (P39).*

In our survey, we asked the participants to rate their wearing frequency in several major places, and the results are presented in Table 8. The survey results showed that wearing frequency was lowest when they were at their home/dorm (2.83 on a 5-point Likert scale, where 1 is *Never wear* and 5 is *Always wear*). Another category with lower frequency was 'gym and sports field.' This is partly because some participants did not want to wear their watches during workout periods. The statistical tests with repeated measures ANOVA showed that these four conditions were significantly different ($F = 21.95, p < 0.001$). Post-hoc pairwise comparison results with Bonferroni correction showed that all pairs were significant ($p < 0.001$), except that there was no significant difference between work/class and restaurant/cafe ($p > 1.0$).

*4.2.2 Diurnal Wearing Patterns.* In Section 3.3, we reported the diurnal wearing patterns and showed that we uncovered three groups of users, namely work-hour wearers, active-hour wearers, and all-day wearers. In our survey, we further divided diurnal activity into the following time slots: (1) After waking up–Before going to work; (2) After going to work–Before leaving work; (3) After leaving work–Before going to bed; and (4) After going to bed–Before waking up. Table 9 shows that 'After going to work–Before leaving work' had an exceptionally high score of 4.43 (SD=1.02), and the other time slots had relatively lower scores. The statistical tests with repeated measures ANOVA showed that these four conditions were significantly different ($F = 42.50, p < 0.001$). Post-hoc pairwise comparison results with Bonferroni correction showed that all pairs were different ($p < 0.001$), except that there was no significant difference between 'After waking up–Before going to work' and 'After leaving work–Before going to bed' ($p > 1.0$).

The participants in the *active-hour wearer* group seldom wore their smartwatches while sleeping. They commented that the discomfort of wearing the watch while sleeping and the need to charge battery are the two

| Time Slot | Wearing Frequency Score |
|---|---|
| After waking up – Before going to work | 2.74 (SD=1.55) |
| After going to work – Before leaving work | 4.43 (SD=1.02) |
| After leaving work – Before going to bed | 2.53 (SD=1.49) |
| After going to bed – Before waking up | 1.38 (SD=0.90) |

Table 9. Wearing frequency rating by time slot

major reasons for low wearing frequency during the night. P14 commented, *"I wear loose clothes when I go to bed, so I think it would be uncomfortable to wear a smartwatch when I sleep. Also, I have to charge the battery" (P14).* Although users may want to track their sleep patterns, discomfort hinders users from doing so, as P28 said, *"I kind of want to track my sleeping patterns, but I don't want to wear the watch in bed. ... Also, if I track my sleeping patterns, I have no time to charge my watch" (P28).* The Apple Watch does not provide default sleep tracking applications, and many participants did not know whether they could install third party apps for sleep tracking.

*4.2.3 Take-off Events.* The measurement analysis results of the survey participants (n=47) showed that the number of take-off events per day was estimated to be 3.23 on average (SD=1.42). In the survey, we asked about take-off events with the following question: "How many times in a day do you take the smartwatch off?" The average response was 2.06 times a day (SD=1.42). The result from the measurement study is slightly higher than the result from survey (3.23 vs. 2.06) ($F = 14.73, p < 0.001$). In our survey, we asked users to rate the major reasons for take-off events, i.e., discomfort, concerns about disturbance, and concerns about breakage. The dominant reasons were *discomfort* (n=28) (e.g., "feeling itchy and sweaty", "uncomfortable while doing computer work," and "uncomfortable while playing basketball") and *concerns about breakage* (n=17) (e.g., "to prevent it from getting wet during experiments"). Our participants commented that take-off events often happened unconsciously, and thus, we posit that users tended to underestimate the number of take-off events. To the question about when they put the watch on again, participants answered that they put their watches on after the activity they were doing was finished, or after the uncomfortable feeling got better. It appears that the average take-off duration of 55.30 minutes from Section 3.3 is a reasonable estimate. Note that we think that the impact of false positives on estimating take-off events is minimal because take-off events are rare, and classification is highly accurate (over 97% in everyday scenarios).

### 4.3 Findings: Factors Affecting Wearing/Taking-Off Behaviors

We provide major factors affecting wearing/taking-off behaviors such as the need of an immediate response, engagement of in social activities, availability of personal workspace, need for multitasking, smartwatch charging, aesthetics, daily activity tracking, and exercise tracking. Overall, we find that the factors affecting wearing behaviors are contextual and nuanced.

*4.3.1 Need for an Immediate Response.* One of the most frequently used features of a smartwatch is to check notifications from the phone, such as incoming calls, messages, and emails. Those participants who wear their watches during the day time commented that smartwatches allowed them to respond immediately, particularly in their work context, as P41 commented, *"I used to miss many calls from other people because I do not check my phone regularly. The Apple Watch helps me in this aspect, so I use it" (P41).* Furthermore, the watch not only helped users to detect incoming notifications, it also gave them the ability to quickly distinguish which notifications are of importance. Users can quickly check from whom the message came and what kind of information it contains with the watch. Before using the smartwatch, users commented that checking notifications with their phones often disturbed their work since it takes more time and effort. Furthermore, checking notifications on the phone was also followed by additional unnecessary smartphone usage (e.g., checking their SNS). P44 commented, *"When I only used my phone, I constantly looked at my phone when I was waiting for a message to be able to respond. But now, I can glance at my watch and check who sent the message, so there is no need to look at my phone. When I look at my smartphone, I end up doing other things besides responding to the message. I can't do those unnecessary activities on the smartwatch, so I like it. It enhanced my working efficiency a lot." (P44)*

Besides incoming contacts from others, users can set alarms to remind themselves to take an action. Examples of these include alarms set with calendar applications before important events, timer alarms for users who do biological or chemical experiments, and timers for the end time of a laundry machine.

However, there are situations, such as non-work hours and weekends, in which users do not have to respond quickly. Our participants explained that the notification function of smartwatches was the reason for deciding not to wear the watch in these situations, because notification alarms made them feel they were still connected to their workplace. In these cases, users intentionally do not wear their watches, as P6 commented, *"This is a watch that alerts me as I have said before. I want to feel that on weekends, I am off from work" (P6).*

*4.3.2 Engagement in Social Activities.* Wearing a smartwatch was useful when our participants were engaged in social activities (e.g., when hanging out with friends, during meetings, and while tutoring) during which smartphone interactions are considered less socially appropriate. P17 said, *"I check almost all my notifications with the Apple Watch. ... While I am in class or in a meeting I always check notifications with this, and also during tutoring" (P17).* In this case, the smartwatch allowed users to quickly glance at incoming notifications, which is less disturbing to other people. One participant even told us about a little trick for using the smartwatch when she is in social situations: *"The watch feels more useful when I am with others. If I look at my smartphone, I feel rude as it seems like I am not listening to others. But when I'm wearing the watch, I can glance and check a notification while pretending I am just checking time" (P39).* Nonetheless, our participants said that vibrations from smartwatches is still disturbing during social activities, but they felt less disturbed compared with smartphone notifications. One major reason is that it is easy to ignore notifications from the smartwatch if they have to focus on social activities.

*4.3.3 Availability of Personal Workspace.* Availability of personal workspace was closely related to take-off behaviors. When users are in a place in which they have their own personal work area, they can take the watch off when they feel uncomfortable. For example, P22 and P39 sometimes took off their watches when they were working in their labs (both of them were mostly working with computers), and P14 took his watch off during class. For undergraduate students, classrooms can be also viewed as a place with a personal workspace since each student has their own desk area, on which they can place their books and pencils. Our participants did not comment on take-off behaviors outside of personal workspaces. In social settings, we expect that unlike smartphones, it is less likely to take off their smartwatches and place them into shared spaces (e.g., a coffee table).

*4.3.4 Need for Multitasking.* Users often wear their smartwatch in situations in which users need to multitasking. One of the main features of a smartwatch as a wearable device is that it can be worn on the wrist. Users can easily manipulate a smartwatch, e.g., for simple checking and browsing. This aspect is especially useful when users are doing activities in which both hands are fully engaged, and significant attention is required; e.g., driving a car, riding a bicycle or conducting an experiment: *"I always used to put my phone inside a pocket when I ride a bicycle. When I got an alarm from my phone, I used to stop my bike and take out the phone to check the notification. But now, I can quickly check notifications and find out whether I got a phone call or a message through the watch while I am in the middle of riding" (P39). "I wear latex gloves during experiments, and it is prohibited to take phones inside the laboratory. ... I can check the time quickly during experiments and also find out who is calling and answer right away when I get a phone call" (P44).*

*4.3.5 Smartwatch Charging.* Short battery duration and the inconvenience of charging have negative effects on smartwatch wearing. Participants said the battery of the Apple Watch lasts only about one and a half days, so they have to charge the device almost every day. Due to the short battery life, users tended to take off the watch and charge it whenever they stayed in a place such as their dormitory for a long time—our participants mostly placed their charging stations at home. The participants sometimes forgot to wear the watch in the morning after having placed it in the charging station the night before. Unlike smartphones, smartwatches have unique charging stations, and it seems like carrying charging stations is burdensome. Our participants often forget to take the charging station with them when traveling, such as going to their parents' house. Furthermore, some participants even commented that they decided not to wear the watch because it is bothersome to carry a separate charging station: *"When I go somewhere else, if I wear the smartwatch, I also have to take the charger, and it becomes another thing to carry. It is not heavy, but I end up thinking, 'I'll just leave it here' " (P6). "When I go home, my brother also uses an iPhone, so I can borrow his iPhone charger even if I don't bring mine. But for the Apple Watch, I have to take the charger separately, and the wire is kind of long, so I often forget to take it with me" (P44).*

*4.3.6 Aesthetics.* Prior studies about smartwatch usage behaviors reported that the aesthetics of a smartwatch is an important factor in selecting and wearing the watch [10, 11, 19, 32, 48]. Some participants in our study opined similar comments. They were using the smartwatch as a substitute for a conventional watch, but they were not fully satisfied with the design of the Apple Watch because its shape and look were significantly different from conventional wrist watches. They wanted to have as similar a feeling to a conventional watch as possible. However, we uncovered slightly different views towards smartwatch design compared to findings in previous studies. Although the Apple Watch did not meet users' expectations as a substitute for a conventional watch, many of the participants were satisfied with the design of the Apple Watch *as a smartwatch.* They concurred that when they wear formal or dressy outfits, the design of the watch does not seem to blend well. Unlike conventional wrist watches whose value as an accessory is deemed to be quite important, we found that our participants considered smartwatch to be *wrist-worn electronic devices* rather than fashion accessories. In general, they were quite *generous* about the aesthetics of the Apple Watch as a wristwatch. P6 and P10 commented: *"I sometimes felt that my smartwatch doesn't blend well with my overall outfits, but it is not relevant since it is an Apple Watch. ... This is a device, not a watch for aesthetics, so it is okay even if the design is not so pretty" (P6); "I feel that it does not match well with my clothes, but I don't care about that. The design of a smartwatch is different from an analog watch. ... This design looks like a smartwatch to everyone, so I don't care about its design too much" (P10).*

*4.3.7 Daily Activity Tracking.* The behavior of people using the daily activity tracking function was similar to that found in related work [28, 29, 47]. Users regularly checked their daily activity (e.g., step counts and calories consumed) and got motivated to move their body more. Sometimes they were just satisfied to be getting information about themselves. Some people participated more actively with daily tracking by following *stand up* or *breathe* alarms from the watch. Users said that occasional smartwatch alarms informing about their physical movement status helped them to refresh themselves during long work hours: *"When I stay in my lab I always sit in front of my desk. I feel the need to get up and move my body, but often don't get up until my back hurts. I get to stand up more often with the Apple Watch before my body gets too stressed out" (P44).* However, some of the participants were not interested in the movement information from the smartwatch. They felt the alarms from the daily activity tracking application disturbed their daily work and became confused with other useful notifications (e.g., emails or messengers). Many of these people decided to turn off the daily tracking function in the watch.

*4.3.8 Exercise Tracking.* For detailed exercise tracking, some users used the native *Workout* app on the Apple Watch or other third-party applications (e.g., Nike Run). However, a majority of users simply preferred the default *Activity* tracking app on the Apple Watch. One reason for this was that our participants found it cumbersome to remember to turn on the *Workout* app. Another reason was that the *Workout* app only supports a limited number of exercises for tracking, as P44 commented: *"I usually do weight training, but I couldn't find out how to set the Workout app for weight training, so I haven't gotten a chance to track my activity during weight training" (P44). "I would like the Workout app to also support minor exercises, like skiing." (P23)*. Moreover, activity tracking was not able to capture movement amounts properly, as P16 commented, *"The exercise I regularly do is yoga. I tried wearing it, but it doesn't seem to track my activity well. Yoga movements are slow and still, so I think the watch doesn't recognize it as an exercise" (P16)*.

While performing intense exercise activities, bulky form factors and concerns of breakage/hurting are the major hindrances of wearing smartwatches. The design of a smartwatch is different from the design of smart bands that mainly focus on activity tracking. A smartwatch has a larger display and thicker body as it supports more functionality and displays more graphical information to users. Due to its large form factor, users felt uncomfortable wearing it while exercising: *"It feels like an ordinary watch during daily activities, but when I exercise, I feel uncomfortable when I bend at my wrist because of the thick shape of the smartwatch" (P44)*. In addition, people are worried about breaking the watch display when they engage in intense workouts, like basketball or soccer. Another problem is that others may get hurt if they are hit by the watch: *"Exercises like ping-pong with a net between players are okay, but soccer or basketball, those types of exercises have a lot of bumps and crashes. So I don't wear the smartwatch when I play game like that and just plug it in the charger" (P10)*.

## 5 DISCUSSION

### 5.1 Summary of Finding and Relevance with Prior Studies

Our work is the first longitudinal measurement study of analyzing smartwatch wearing behaviors by collecting activity tracking data (step counts and heart rates) from 50 Apple Watch users over 200 days. We adopted a mixed method by quantitatively analyzing temporal wearing patterns and qualitatively investigating reasons and factors related to wearing behaviors.

Toward this goal, we built an activity tracker data collection platform. Since the Apple Watch does not provide any information about wearing status, we propose a novel approach of inferring wearing status only based on the activity tracker data, namely heart rates and step counts. After collecting a ground truth dataset, we analyzed the characteristics of activity tracker data under various conditions and found that heart rate or step count data cannot be directly used for wearing state inference. We extracted various features from the raw data and evaluated machine learning models. The results showed that we can achieve over 91% accuracy with a detection time granularity of eight minutes. To validate our model in naturalistic wearing scenarios, we also collected a week-long dataset from four users and showed that our model is highly accurate, achieving average classification accuracy of 97%. Overall, our approach can be applicable to the other conventional health trackers that do not have a dedicated sensor for tracking wearing status.

Our analysis results (n=50) showed that the mean daily wearing time were 10.18 hours (SD=3.51), and the mean number of take-off events per day was 3.17 (SD=1.11). However, our survey results (n=47) showed that subjective reports were less accurate. Participants overestimated the mean wearing duration as 13.38 hours (SD=3.65), whereas they underestimated the mean number of take-off events per day as 2.06 (SD=1.42). Note that self-report is frequently used method when researchers try to understand the using behaviors of users, especially when the subject of study is in an early adoption stage. This method is useful as researchers could get approximate numbers related to using behaviors without making measuring technologies or experiments. However, our results indicate that users tend to inaccurately report their usage behaviors, and thus, self-report data should be carefully interpreted.

We identified three distinct groups of usage patterns, namely work-hour wearers (58%), active-hour wearers (30%), and all-day wearers (12%). The majority of the participants belong to the work-hour and active-hour wearer groups. These wearing patterns are somewhat different from those of traditional watches [32] in that users tend to wear smartwatches longer hours. We hypothesize that such behaviors may be due to the usefulness of smartwatches such as notification checking and activity tracking, which would be true for 42% of the participants (i.e., active-hour and all-day wearers).

Usage density of a user is fairly high with its average of 88.7%, which means that on average our participants wore their watches for 6.21 days per week. Users' average daily wearing hours per week were fairly consistent over time with the median wearing hours per day ranging in between 8.34 and 13.87. Users tended to have short break duration of several days (mostly a few days of not wearing) and there were significant individual variations of break and streak length distributions. Manual examination of break length distribution per user revealed two groups of users: power users, whose median break duration is just 1 day (n=19), and casual users, whose median break duration is greater than 1 day (n=31). Despite causal usage, a majority of casual users had low casualness (n=27) in that they consistently wore their smartwatches over the entire period, but they may not wear their watches everyday and may have longer breaks sometimes. These kinds of consistent long term usage patterns were also observed in their activity tracker study by Meyer et al. [34, 35]. Nonetheless, we found that the level of engagement is higher in smartwatches than that in activity trackers because smartwatches are used for multiple purposes such as message checking and activity tracking.

Our qualitative results showed that daily routines and places are closely related to the wearing behaviors, and home has the lowest score of wearing when compared with other places such as workplace and cafeteria. The major reasons for take-off events were discomfort and concerns of breakage. Wearing behaviors of smartwatch is not affected by a single major cause, but a result of multiple aspects together: the need for an immediate response, engagement in social activities, availability of personal workspace, need for multitasking, smartwatch charging, aesthetic concerns, daily activity tracking, and exercise tracking. Overall, we find that the factors affecting wearing behaviors are contextual and nuanced. As presented in prior studies [9, 42, 48], the major facilitator was their usefulness under various contexts such as checking notifications and tracking physical activities. Affording physical presence as a wrist-worn wearable (related to materiality) and micro-interactions facilitates convenient/fast information access and helps to sustain wearing behaviors [3, 42].

## 5.2 Design Implications

There are a number of issues that have negative effects on wearing behaviors such as burdens of immediate responses, disturbance in social contexts, limited/inaccurate activity tracking, concerns of breakage, battery charging, and aesthetic concerns. Some of these findings were consistent with the prior studies [10, 13, 37]. In the following, we discuss various strategies of improving wearbility of smartwatches and leveraging smartwatches for behavioral changes

*Nudging Smartwatch Wearing*: Our results showed that short breaks are commonly observed throughout days of week (beyond weekends). Interview results revealed that such breaks were mostly because they forgot to take their watches on the charging stations. One way of improving wearability is to provide a context-aware reminder in their smartphones. Bluetooth connectivity can be used to remind whether their watches are nearby or not or place-based reminding can be leveraged (e.g., detecting place departure based on WiFi fingerprinting). Furthermore, we can leverage various visualization techniques to improve awareness of wearing behaviors, as Epstein et al. [14] suggested.

*Managing Interruptions in Smartwatches*: Smartwatches support micro-interactions that take only several seconds to initiate and complete such as notification checking [3]. Micro-interactions with smartwatches were generally considered as less disrupting than those with smartphones, but we found that they can be disruptive in some contexts (e.g., stress of constant connectivity at home). One way of mitigating such smartwatch interruptions is to enable context-aware interruption management [40, 41].

*Dealing with Charging Inconvenience*: Our participants habitually charged their watches everyday (mostly at home) and tended not to carry their charging stations due to inconvenience. The smartwatch chargers differ in shape from the smartphone chargers. It is hard to find an alternative power source if users do not carry the smartwatch charger. In reality, users have diverse functional needs with smartwatches (e.g., time checking, notification reception, activity tracking). Thus, dealing with charging inconvenience is critical because this may degrade wearing behaviors, which could lead to abandonment as in fitness trackers [12, 13].

*Supporting Diverse Activity Tracking*: Some of our participants commented that due to limited activity tracking capabilities they tended not to wear their watches while exercising. The current exercise types provided by activity tracking apps in smartwatches mostly focus on major exercises such as walking and running. Throughout the study, we found that assuming users would be satisfied with tracking the major exercise types is misleading. While recent activity trackers support diverse types of activity tracking, smartwatches currently lack such options. Thus, we can increase the wearability of smartwatches by supporting tracking features of diverse physical activities.

*Leveraging Smartwatches for Behavioral Changes*: Our results showed that smartwatches had higher wearing intensity as opposed to activity trackers due to its support for multiple functions ranging from notification checking to activity tracking, by complementing their smartphones. Due to diverse functional support and smartphone interconnectivity, we expect that smartwatches can be considered as gateway/deficit technologies for training/routinizing new behaviors, as Fritz et al. [16] suggested. Despite possibility of increased interruption, micro-interactions with smartwatches provide immense opportunities for timely delivery of behavioral change related information for prompt *checking/glancing*, which may lead to detailed review/reflection with their smartphones and thereby to durable behavioral changes [17]. Furthermore, smartwatches can be used as ways of *enforcing* certain behaviors by causing discomfort/inconvenience to users [4, 46, 50]; e.g., vibrating watches until the users show desirable behaviors. For example, as one of the ways of mitigating smartphone overuse, unlike conventional lock-based self-limiting mechanisms [23–25], we can nudge users of overuse by vibrating watches until the users stop using their phones.

## 5.3 Limitations

Our study has several limitations about generalizability of the findings because we only tested single smartwatch with college students. Overall, to generalize our findings, there should be comparative studies with other smartwatch devices and diverse user groups.

First of all, we only examined the Apple Watch users, which has the largest market share in the world (45.6% as of Q3 2016). The measurement and survey/interview results could be biased due to particular characteristics the Apple Watch, such as its shape and the high novelty of Apple product users. Thus, temporal wearing behaviors and factors affecting wearing behaviors could be different in other smartwatches.

Another limitation is that the results might be biased because the participants were college students. The life patterns of college students are distinctively different from those of other groups, and this behavioral pattern could have affected the way our participants wear their smartwatches. Nonetheless, analyzing physical wellbeing of college students had received a lot of attention in recent years. For example, the StudentLife study analyzed social, emotional, and physical health of college students using smartphone-based sensing [52]. The NetHealth study analyzed the longitudinal Fitbit dataset of college freshmen to understand their wellbeing in campus [43]. Note that our work investigated the college students as a focusing lens on emerging wearable technologies. Younger generations are likely to be early adoptors of novel technologies. Furthermore, college students are considered to be vulnerable to various physical and mental well-being because of their developmental dynamics and relative independence from social roles and expectations [21, 43, 52].

There is also a concern about our data collection method, because we freely distributed smartwatches to the study participants. In our interview, some participants mentioned sense of obligations and burdens for wearing. This kind of reciprocal usage motivation may affect users to give longer wearing hours of smartwatch. In the early stage of technology adoption, however, such data collection method is often used in the ubicomp literature [26, 49]. Furthermore, our method is very similar to recent large-scale measurement studies of wearable trackers where for data collection, researchers freely distributed activity trackers to specific interest groups (e.g., patients, college students) [34, 35, 43].

In our study, we used heart rate and step count data to infer wearing behaviors, because the Apple Watch does not provide any API for accessing such information. Furthermore, to our knowledge, none of the current smartwatches in the market provide any APIs to check wearing status. Instead of using heart rate samples and step count data, processing raw sensor signals from heart rate sensors and skin conductance sensors allows us to build classification models for inferring wearing status. For example, highly sampled raw heart beat signals may differ significantly depending on wearing status. Exploring energy efficient and accurate sensing mechanisms for detecting wearing status would be an interesting direction for future work.

## 6 CONCLUSION

As representative wearable devices, we studied wearing behaviors of smartwatches, which is critical for designing wearable devices for various purposes. Towards this goal, we analyzed temporal wearing patterns of smartwatches by collecting a longitudinal activity dataset from 50 smartwatch users for just over 200 days. Furthermore, we investigated the reasons for wearing behaviors qualitatively with a survey and interviews. Through the study, we identified three groups of smartwatch users based on their diurnal wearing patterns such as work-hour wearers, active-hour wearers, and all-day wearers. Temporal behavior analysis showed that usage density of a user is fairly high and consistent with only short break duration, possibly due to high engagement of smartwatch features. Our qualitative study results showed that the factors affecting wearing behaviors are contextual, nuanced, and multifaceted, including micro-interaction needs in daily routines, smartwatch charging, aesthetic concerns, and activity/exercise tracking accuracy. We discussed several practical design considerations for improving wearability of smartwatches and leveraging smartwatches for delivering behavioral intervention. There are several directions for future work. First of all, for generalizability there should be follow-up measurement studies with different user groups and trackers. Second, fine-grained temporal and spatial analysis of wearing behaviors will allow us to quantify how various contextual factors influence wearing behaviors. Third, collecting and analyzing multi-device usage data will provide better insights into wearing behaviors of smartwatches in the ecology of ubiquitous computing.

## ACKNOWLEDGMENTS

This work was partly supported by KAIST KI for Health Science and Technology, and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korea government (MSIP) (No. NRF-2015R1D1A1A01059497). The corresponding authors of this work are Uichin Lee (uclee@kaist.ac.kr) and Heepyung Kim (heepkim@kaist.ac.kr).

## REFERENCES

[1] Apple, Inc. 2016. Our Heart Rate. What It Means, and Where on Apple Watch You'll Find it. (2016). https://support .apple.com/en-ca/HT204666.

[2] Apple, Inc. 2016. UIApplicationBackgroundFetchIntervalMinimum. (2016). https://developer.apple.com/reference/uikit/uiapplicationbackgroundfetchintervalminimum.

[3] Daniel L. Ashbrook. 2010. *Enabling Mobile Microinteractions.* Ph.D. Dissertation. Atlanta, GA, USA. Advisor(s) Starner, Thad E. AAI3414437.

[4] Steve Benford, Chris Greenhalgh, Gabriella Giannachi, Brendan Walker, Joe Marshall, and Tom Rodden. 2013. Uncomfortable User Experience. *Commun. ACM* 56, 9 (Sept. 2013), 66–73. DOI: http://dx.doi.org/10.1145/2500468.2500889

[5] Yannick Bernaerts, Matthias Druwé, Sebastiaan Steensels, Jo Vermeulen, and Johannes Schöning. 2014. The Office Smartwatch: Development and Design of a Smartwatch App to Digitally Augment Interactions in an Office Environment. In *Proceedings of the 2014 Companion Publication on Designing Interactive Systems (DIS Companion '14)*. ACM, New York, NY, USA, 41–44. DOI: http://dx.doi.org/10.1145/2598784.2602777

[6] Hugh Beyer and Karen Holtzblatt. 1997. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann.

[7] Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A Tutorial on Human Activity Recognition Using Body-worn Inertial Sensors. *ACM Comput. Surv.* 46, 3, Article 33 (Jan. 2014), 33 pages. DOI: http://dx.doi.org/10.1145/2499621

[8] Canalys 2016. Media Alert: Smart Watch Market Grows 60% in Q3 2016 as Apple Ships 2.8 Million Units. (2016). https://www.canalys.com/newsroom/media-alert-smart-watch-market-grows-60-q3-2016-apple-ships-28-million-units.

[9] Marta E. Cecchinato, Anna L. Cox, and Jon Bird. 2015. Smartwatches: The Good, the Bad and the Ugly?. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 2133–2138. DOI: http://dx.doi.org/10.1145/2702613.2732837

[10] Jaewon Choi and Seongcheol Kim. 2016. Is the Smartwatch an IT Product or a Fashion Product? A Study on Factors Affecting the Intention to Use Smartwatches. *Computers in Human Behavior* 63 (2016), 777 – 786. DOI: http://dx.doi.org/10.1016/j.chb.2016.06.007

[11] Stephanie Hui-Wen Chuah, Philipp A. Rauschnabel, Nina Krey, Bang Nguyen, Thurasamy Ramayah, and Shwetak Lade. 2016. Wearable Technologies: The Role of Usefulness and Visibility in Smartwatch Adoption. *Computers in Human Behavior* 65 (2016), 276 – 284. DOI: http://dx.doi.org/10.1016/j.chb.2016.07.047

[12] James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. 2015. No Longer Wearing: Investigating the Abandonment of Personal Health-tracking Technologies on Craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 647–658. DOI: http://dx.doi.org/10.1145/2750858.2807554

[13] Daniel A. Epstein, Monica Caraway, Chuck Johnston, An Ping, James Fogarty, and Sean A. Munson. 2016. Beyond Abandonment to Next Steps: Understanding and Designing for Life After Personal Informatics Tool Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1109–1113. DOI: http://dx.doi.org/10.1145/2858036.2858045

[14] Daniel A. Epstein, Jennifer H. Kang, Laura R. Pina, James Fogarty, and Sean A. Munson. 2016. Reconsidering the Device in the Drawer: Lapses As a Design Opportunity in Personal Informatics. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 829–840. DOI: http://dx.doi.org/10.1145/2971648.2971656

[15] Alexander W. Ernst and Claus-Peter H. Ernst. 2016. *Success Comes to Those Who Are Successful: The Influence of Past Product Expectation Confirmation on Smartwatch Usage*. Springer International Publishing, Cham. 49–58 pages. DOI: http://dx.doi.org/10.1007/978-3-319-30376-5_5

[16] Thomas Fritz, Elaine M. Huang, Gail C. Murphy, and Thomas Zimmermann. 2014. Persuasive Technology in the Real World: A Study of Long-term Use of Activity Sensing Devices for Fitness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 487–496. DOI: http://dx.doi.org/10.1145/2556288.2557383

[17] Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. 2015. How Do We Engage with Activity Trackers?: A Longitudinal Study of Habito. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 1305–1316. DOI: http://dx.doi.org/10.1145/2750858.2804290

[18] Haibo He and Edwardo A. Garcia. 2009. Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering* 21, 9 (Sept 2009), 1263–1284. DOI: http://dx.doi.org/10.1109/TKDE.2008.239

[19] Yoonhyuk Jung, Seongcheol Kim, and Boreum Choi. 2016. Consumer Valuation of the Wearables: The Case of Smartwatches. *Computers in Human Behavior* 63 (2016), 899 – 905. DOI: http://dx.doi.org/10.1016/j.chb.2016.06.040

[20] Haik Kalantarian and Majid Sarrafzadeh. 2015. Audio-based Detection and Evaluation of Eating Behavior using the Smartwatch Platform. *Computers in Biology and Medicine* 65 (2015), 1 – 9. DOI: http://dx.doi.org/10.1016/j.compbiomed.2015.07.013

[21] Jonathan J. Kandell. 1998. Internet Addiction on Campus: The Vulnerability of College Students. *CyberPsychology & Behavior* 1, 1 (1998).

[22] Evangelos Karapanos, Rben Gouveia, Marc Hassenzahl, and Jodi Forlizzi. 2016. Wellbeing in the Making: Peoples' Experiences with Wearable Activity Trackers. *Psychology of Well-Being* 6, 4 (2016).

[23] Minsam Ko, Seungwoo Choi, Subin Yang, Joonwon Lee, and Uichin Lee. 2015. FamiLync: Facilitating Participatory Parental Mediation of Adolescents' smartphone use. In *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 867–878.

[24] Minsam Ko, Seungwoo Choi, Koji Yatani, and Uichin Lee. 2016. Lock n'LoL: Group-based Limiting Assistance App to Mitigate Smartphone Distractions in Group Activities. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 998–1010.

[25] Minsam Ko, Subin Yang, Joonwon Lee, Christian Heizmann, Jinyoung Jeong, Uichin Lee, Daehee Shin, Koji Yatani, Junehwa Song, and Kyong-Mee Chung. 2015. NUGU: A Group-based Intervention App for Improving Self-regulation of Limiting Smartphone Use. In *Proceedings of the Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1235–1245.

[26] Amanda Lazar, Christian Koehler, Joshua Tanenbaum, and David H. Nguyen. 2015. Why We Use and Abandon Smart Devices. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 635–646. DOI: http://dx.doi.org/10.1145/2750858.2804288

[27] Uichin Lee, Joonwon Lee, Minsam Ko, Changhun Lee, Yuhwan Kim, Subin Yang, Koji Yatani, Gahgene Gweon, Kyong-Mee Chung, and Junehwa Song. 2014. Hooked on Smartphones: An Exploratory Study on Smartphone Overuse Among College Students. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2327–2336. DOI: http://dx.doi.org/10.1145/2556288.2557366

[28] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A Stage-based Model of Personal Informatics Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 557–566. DOI: http://dx.doi.org/10.1145/1753326.1753409

[29] Ian Li, Anind K. Dey, and Jodi Forlizzi. 2011. Understanding My Data, Myself: Supporting Self-reflection with Ubicomp Technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11)*. ACM, New York, NY, USA, 405–414. DOI: http://dx.doi.org/10.1145/2030112.2030166

[30] Hong Lu, Wei Pan, Nicholas D. Lane, Tanzeem Choudhury, and Andrew T. Campbell. 2009. SoundSense: Scalable Sound Sensing for People-centric Applications on Mobile Phones. In *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services (MobiSys '09)*. ACM, New York, NY, USA, 165–178. DOI: http://dx.doi.org/10.1145/1555816.1555834

[31] Jay Lundell and Corrie Bates. 2016. *Understanding User Experience Journeys for a Smart Watch Device*. Springer International Publishing, Cham. 424–433 pages. DOI: http://dx.doi.org/10.1007/978-3-319-39399-5_40

[32] Kent Lyons. 2015. What Can a Dumb Watch Teach a Smartwatch?: Informing the Design of Smartwatches. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers (ISWC '15)*. ACM, New York, NY, USA, 3–10. DOI: http://dx.doi.org/10.1145/2802083.2802084

[33] Kent Lyons. 2016. Smartwatch Innovation: Exploring a Watch-First Model. *IEEE Pervasive Computing* 15, 1 (Jan 2016), 10–13. DOI: http://dx.doi.org/10.1109/MPRV.2016.21

[34] Jochen Meyer, Jochen Schnauber, Wilko Heuten, Harm Wienbergen, Rainer Hambrecht, Hans-JRürgen Appelrath, and Susanne Boll. 2016. Exploring Longitudinal Use of Activity Trackers. In *2016 IEEE International Conference on Healthcare Informatics (ICHI)*. 198–206. DOI: http://dx.doi.org/10.1109/ICHI.2016.29

[35] Jochen Meyer, Merlin Wasmann, Wilko Heuten, Abdallah El Ali, and Susanne C.J. Boll. 2017. Identification and Classification of Usage Patterns in Long-Term Activity Tracking. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 667–678. DOI: http://dx.doi.org/10.1145/3025453.3025690

[36] Chulhong Min, Seungwoo Kang, Chungkuk Yoo, Jeehoon Cha, Sangwon Choi, Younghan Oh, and Junehwa Song. 2015. Exploring Current Practices for Battery Use and Management of Smartwatches. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers (ISWC '15)*. ACM, New York, NY, USA, 11–18. DOI: http://dx.doi.org/10.1145/2802083.2802085

[37] Vivian Genaro Motti and Kelly Caine. 2016. Smart Wearables or Dumb Wearables?: Understanding How Context Impacts the UX in Wrist Worn Interaction. In *Proceedings of the 34th ACM International Conference on the Design of Communication (SIGDOC '16)*. ACM, New York, NY, USA, Article 10, 10 pages. DOI: http://dx.doi.org/10.1145/2987592.2987606

[38] Ian Oakley, DoYoung Lee, MD. Rasel Islam, and Augusto Esteves. 2015. Beats: Tapping Gestures for Smart Watches. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1237–1246. DOI: http://dx.doi.org/10.1145/2702123.2702226

[39] Jeungmin Oh and Uichin Lee. 2015. Exploring UX Issues in Quantified Self Technologies. In *Eighth International Conference on Mobile Computing and Ubiquitous Networkin (ICMU '15)*.

[40] Chunjong Park, Junsung Lim, Juho Kim, Sung-Ju Lee, and Dongman Lee. 2017. "Don't Bother Me. I'm Socializing!": Breakpoint-Based Smartphone Notification System. In *Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '17)*. ACM, New York, NY, USA.

[41] Veljko Pejovic and Mirco Musolesi. 2014. InterruptMe: Designing Intelligent Prompting Mechanisms for Pervasive Applications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 897–908. DOI: http://dx.doi.org/10.1145/2632048.2632062

[42] Stefania Pizza, Barry Brown, Donald McMillan, and Airi Lampinen. 2016. Smartwatch in Vivo. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 5456–5469. DOI: http://dx.doi.org/10.1145/2858036.2858522

[43] Rachael Purta, Stephen Mattingly, Lixing Song, Omar Lizardo, David Hachen, Christian Poellabauer, and Aaron Striegel. 2016. Experiences Measuring Sleep and Physical Activity Patterns Across a Large College Cohort with Fitbits. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers (ISWC '16)*. ACM, New York, NY, USA, 28–35. DOI: http://dx.doi.org/10.1145/2971763.2971767

[44] Reza Rawassizadeh, Blaine A. Price, and Marian Petre. 2014. Wearables: Has the Age of Smartwatches Finally Arrived? *Commun. ACM* 58, 1 (Dec. 2014), 45–47. DOI: http://dx.doi.org/10.1145/2629633

[45] Sasank Reddy, Min Mun, Jeff Burke, Deborah Estrin, Mark Hansen, and Mani Srivastava. 2010. Using Mobile Phones to Determine Transportation Modes. *ACM Trans. Sen. Netw.* 6, 2, Article 13 (March 2010), 27 pages. DOI: http://dx.doi.org/10.1145/1689239.1689243

[46] Jun Rekimoto and Hitomi Tsujita. 2014. Inconvenient Interactions: An Alternative Interaction Design Approach to Enrich Our Daily Activities. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces (AVI '14)*. ACM, New York, NY, USA, 225–228. DOI:http://dx.doi.org/10.1145/2598153.2598196

[47] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers Chalmers. 2014. Personal Tracking As Lived Informatics. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1163–1172. DOI:http://dx.doi.org/10.1145/2556288.2557039

[48] Steven Schirra and Frank R. Bentley. 2015. "It's Kind of Like an Extra Screen for My Phone": Understanding Everyday Uses of Consumer Smart Watches. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 2151–2156. DOI:http://dx.doi.org/10.1145/2702613.2732931

[49] Patrick C. Shih, Kyungsik Han, Erika Shehan Poole, Mary Beth Rosson, and John M. Carroll. 2015. Why We Use and Abandon Smart Devices. In *Proceedings of iConference*.

[50] Hitomi Tsujita and Jun Rekimoto. 2011. Smiling Makes Us Happier: Enhancing Positive Mood and Communication with Smile-encouraging Digital Appliances. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11)*. ACM, New York, NY, USA, 1–10. DOI:http://dx.doi.org/10.1145/2030112.2030114

[51] Ulrike von Luxburg. 2007. A Tutorial on Spectral Clustering. *Statistics and Computing* 17, 4 (2007), 395–416. DOI:http://dx.doi.org/10.1007/s11222-007-9033-z

[52] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. 2014. StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. ACM, New York, NY, USA, 3–14. DOI:http://dx.doi.org/10.1145/2632048.2632054

[53] Dominik Weber, Alexandra Voit, Philipp Kratzer, and Niels Henze. 2016. In-situ Investigation of Notifications in Multi-device Environments. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 1259–1264. DOI:http://dx.doi.org/10.1145/2971648.2971732

[54] Cheng Xu and Kent Lyons. 2015. Shimmering Smartwatches: Exploring the Smartwatch Design Space. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '15)*. ACM, New York, NY, USA, 69–76. DOI:http://dx.doi.org/10.1145/2677199.2680599

Received February 2017; revised May 2017; accepted July 2017.