Appendix A

Technical Information to Facilitate Key Design Decisions

Selecting a Smartphone Device and Sensing Application

At the time of this writing, Android devices provide more flexibility than do iOS devices with respect to collecting data in smartphone-sensing studies. The flexibility of the Android device stems from two features. First, Android devices allow third-party applications to collect data from many of the phone's sensors and operating system logs (e.g., application usage logs, call and SMS logs, battery and charging logs). Second, Android devices allow third-party applications to run continuously, collecting data as part of the background processes of the phone, even when the application is not being used.

In contrast, the iOS-based iPhone allows third-party applications to collect data only from a limited number of sensors, and does not allow access to the operating-system logs. The iPhone also does not currently allow third-party applications to run in the background processes on the phone (see Miluzzo et al., 2008 for an in-depth evaluation of mobile sensing on the iPhone). Instead, current mobile sensing applications designed for iPhones typically can collect sensor data only when the application is open and running in the foreground of the phone screen. One caveat to this restriction is that iPhones permit the collection of periodic location data even when the application is in the background, by capturing significant changes in location that are picked up by a combination of cell towers and GPS. Thus, the decision to use one device over another will affect the types of sensor data collected and the sampling rates (i.e., continuous vs. periodic data collection).

Recent updates to the iPhone OS suggest that iOS devices may permit flexible sensor data collection for third-party applications in the near future. However, given the current limitations to sensor-data collection on the iPhone, we recommend that researchers interested in continuous sampling of many sensors (e.g., sampling every few minutes) use Android devices. iOS devices should be primarily used by researchers interested in periodic sampling from the accelerometer and GPS sensors (e.g., every 15 minutes, or each time the application is opened) or where high numbers of the participant pool have iPhones (e.g., many student samples).

An additional factor to consider when selecting the smartphone device will be whether the researchers will provide the participants with mobile devices, or require participants to use their own mobile devices. Both of these techniques have been successfully used in previous research, and there are benefits and drawbacks to each approach. The benefit of providing participants with mobile devices for use during the study is that the application will technically function the same for all the participants. For example, in the StudentLife study (Wang et al., 2014) participants were loaned an Android phone with the application already installed to use for the duration of the study. This approach resulted in fewer technical problems experienced by participants (e.g., fewer bugs in the application that cause freezing or crashing) because all the participants used the same OS. However, the drawback of providing participants with mobile devices is that they may or may not make it their primary device, which could influence the behavioral estimates computed from the sensor data. For example, if participants are asked to carry around a secondary device, there will likely be instances in which they forget to bring the device with them, potentially underestimating the amount of time they spend engaging in various behaviors (e.g., being physically active or socializing). Figure S1 is reproduced from the StudentLife study (Wang et al., 2014) to illustrate the differences in data collection that may result from participants using their own devices vs. a secondary device.

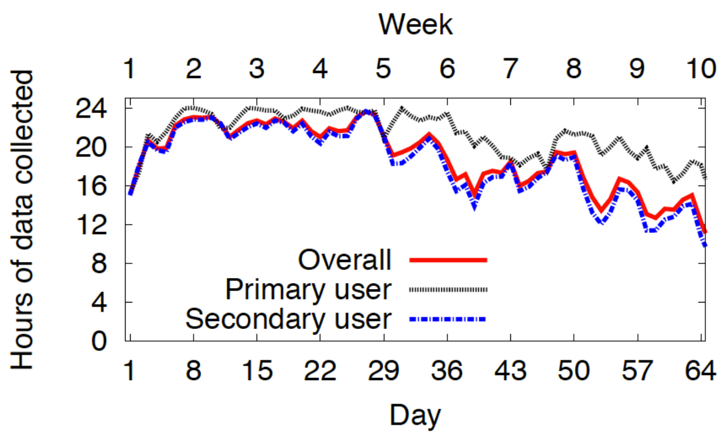

Figure S1. Example of data monitoring technique for visualizing collected sensor data. The graph depicts the average number of hours of sensor data collected from the participants throughout the study duration. Primary users (N = 11) were participants who installed the application to their own smartphones. Secondary users (N = 37) were participants who carried an additional (i.e., secondary) smartphone with the application installed on it. This graph was previously published in the StudentLife study (Wang et al., 2014).

Appendix B

Information about the Logistical Setup of Sensing Systems

Front-end component

The first component of a smartphone sensing system is the frontend, which consists of the smartphone application software that is installed on the participant's phone. The sensing-application software collects data by sampling from a series of sensors, applications, and phone logs. To prolong the battery life of the smartphone and avoid accruing charges on the participant's own data plan, the application may store the collected data on the phone until the phone is charging and/or connected to WiFi. When the phone is connected to WiFi, the application then uploads the data to the portal server (see Back-end description below) using a Secure Sockets Layer (SSL) encryption for added security. The SSL permits the secure transmission of data between the phone and the server. Once the data are uploaded to the server, they remain stored there until they are ready to be extracted for processing and subsequent analysis.

Back-end component

The back-end component runs behind the scenes on a server. There are typically three major features of the back-end component: the portal sever, the participant-manager, and the data storage. The portal sever is the main node of the back-end component because it receives the data from the smartphone front-end and checks it against the participant-manager (which provides user authentication) before storing it in the data storage.

The participant-manager is a user-authentication feature, which contains a list of registered participants for the study and permits researchers to control who participates in the study. When a participants' smartphone attempts to upload his/her data to the portal server, the front-end first sends its participant's authentication (typically in the form of a unique identifier) to the portal sever. The portal sever uses the participant-manager to check if the participant's authentication is valid (i.e., registered for the study). If the participant-manager shows the authentication to be valid, the front-end is permitted to upload the sensor data to the portal server. The portal server then knows who the data belongs to, and parses the uploaded data and saves it in the appropriate location in the data storage.

The data-storage requirements for sensing studies go beyond those of traditional psychological studies because of the continuous data-collection process. Typically, physical servers or cloud servers are purchased to store the sensor data. Physical servers require that researchers estimate the amount of data storage needed prior to data collection. However, it may be difficult to estimate the size of the sensing dataset prior to data collection, particularly for studies using different OSs because inconsistencies in sampling rates (e.g., across devices or OS platforms) can lead to varying amounts of data collected per participant. Moreover, we suspect the size of sensor datasets will increase exponentially as sensing studies include larger, representative samples (e.g., participants recruited world-wide via application stores), collect data over longer periods of time (e.g., for as long as a participant chooses to keep the application installed), and include more data-heavy measures (e.g., high quality video). To illustrate the dataset sizes that can be reached, consider the StudentLife study (Wang et al., 2014) that continuously tracked 48 students using a smartphone application for a ten-week academic term. The total amount of sensor data collected was 52.6 GB, for an average of 1.1 GB per person. Thus, we recommend researchers use cloud servers, such as Amazon's Elastic Compute Cloud, for data storage, especially when the size of the dataset cannot be predicted accurately, because cloud servers provide flexible storage capabilities.

Data-processing component

The third component of a smartphone sensing system is data processing, which typically consists of monitoring the data collection, preparing the sensor data for analysis, and the formal data analysis. Real-time data monitoring allows researchers to determine how long the app is running (collecting data) per day, for each participant. Pre-set thresholds (e.g., less than 15 hours of data collection per day in a study using a continuous sampling rate) can be used to trigger investigations into potential data-collection issues; for instance, a participant may have uninstalled the application or forgotten to turn on their WiFi to allow the data to be uploaded to the portal server. In these instances, the researcher would want to identify the problem and contact the participant (e.g., to find out why they dropped out of the study, remind them to connect to WiFi when possible). Another practice is to calculate how many hours of sensor data (e.g., accelerometer, GPS, Bluetooth) have been collected for each day of the study. Figure S1 presents an example of this data monitoring practice from the data collected in the StudentLife study (Wang et al., 2014). These types of visualizations can help to identify problems in data collection, such as drops in participation that may be due to technical issues or general attrition.