Personalized Implicit Health Monitors

Nicholas Chen, Yun Young Lee, Maurice Rabb, Bruce Schatz
Department of Computer Science
University of Illinois at Urbana-Champaign
{nchen, lee467, m3rabb, schatz}@illinois.edu

ABSTRACT

A person's psychological well-being can be deduced by observing her past, current and future behaviors. However, little effort has been made to qualify, quantify and correlate a person's behavior to her psychological well-being using *non-intrusive* health monitors. This report presents our attempts at using non-intrusive health monitors to observe a person's diet, exercise and sleep patterns to determine possible correlations with her stress levels – a common measure of psychological well-being. Our preliminary study of monitoring three subjects *daily* for a period of seven *continuous* weeks shows that such non-intrusive monitoring yields interesting insights in correlating a person's stress levels. Our study also reveals important design decisions that should be considered in order to reliably and effectively deploy long-term personalized implicit health monitoring systems involving ordinary people in the real world.

Categories and Subject Descriptors

J.3 [**Life and Medical Sciences**]: Health; J.4 [**Social and Behavioral Sciences**]: Psychology; K.4.1 [**Computers and Society**]: Public Policy Issues—*Computer-related health issues*

General Terms

Design, Experimentation, Human Factors, Measurement

Keywords

Health Monitors, Stress

1. INTRODUCTION

The World Health Organization (WHO), in its constitution, defines health as "the state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" [18]. While the healthcare system has made great advances in the treatment of acute diseases, little has been done to help assure physical, mental and social well-being in the population.

Assuring a person's mental and social well-being has always been hard – it is often hard to detect when such an illness occurs. For depression, which affects approximately 14.8 million American adults in a given year [8]. Diagnosing depression requires a complete diagnostic evaluation performed by a doctor or mental health professional.

However, consider some of the symptoms of depression [12] which include:

- Insomnia, early-morning wakefulness or excessive sleeping;
- Overeating or loss of appetite;
- Loss of interest in activities or hobbies once pleasurable.

All those symptoms are easily observable common behavior. A close friend, colleague or family member could easily detect such behavioral changes in someone who might be suffering from depression. While such behavioral changes do not necessarily indicate depression, they strongly suggest the possibility of depression and that profession care may be warranted. As with many illnesses, the earlier depression is detected, the easier it is to treat and prevent.

Using current technology, is there a way to help detect such detrimental behavioral changes easily on an individual basis and suggest visits to the doctor as necessary? In other words, is it possible to create a *personalized implicit health monitoring* system?

Such a system should be *personalized* to observe changes in individual behavior. For example, one person's typical behavior might be to eat a fixed portion at each meal; but when he is depressed he eats more. On the other hand, another person loses her appetite when she is depressed. Each individual has different behaviors that are considered typical. A personalized system will account for each individual's typical behavior and detect deviations from that.

Furthermore, such a systems should also be *implicit.* Instead of directly measuring for depression, the system should monitor for changes in other aspects of behavior. Subtle changes in other aspects of behavior usually manifest before depression itself is detected.

Finally, such a system should rely on automatic *health monitors* whenever possible. Depending on patients to report their behaviors is unreliable. A patient might overstate their symptoms when they are first admitted in order to get treatment faster. They might understate their symptoms when they are trying to end treatment. Health monitors produce more honest and accurate results.

This report presents our prototype implementation using non-intrusive *health monitors* to monitor a subject's behavior under their typical environment. Monitoring normal behavior necessitates monitoring subjects in their normal environments. Requiring subjects to live in specially designed environments, as discussed in Section 7, will inadvertently change their behaviors.

For the purpose of this experiment, instead of depression symptoms, we measure stress-related symptoms. Stress can be viewed as a milder form of anxiety, which itself is a milder form of depression. Thus, it shares many of the common symptoms that depression has but in less pronounced forms. If our implementation works for detecting behavioral changes due to stress, it would be easy to adapt it to detect behavioral changes due to depression as well.

It is easier to observe behavioral changes due to stress instead of depression. Most adults oscillate between periods of high stress and low stress in their daily lives and we can easily observe any behavioral changes. On the other hand, it is neither feasible nor ethical to deliberately induce depression in the subjects for our study.

Our report makes the following contributions toward creating personalized implicit health monitoring systems:

- The design of a system for determining stress from behavior using non-intrusive health monitors. Subjects were able to monitor themselves daily for the period of the study and provided us with seven weeks of data for analysis. They resumed their daily activities and were not required to live in any specially designed environment. As far as we know, this is the first attempt at such a system for monitoring subjects in their typical environment.
- The results from evaluating of our system on three subjects for seven weeks. For future work, we plan to increase the number of subjects and the duration of the study. For this initial study, the three authors acted as the subjects.
- The design of an application that allows the subjects to visualize their behavior and detect any atypical behaviors. A summary of our system is provided in the appendix.

2. SYSTEM DESIGN

In designing a personalized system, there are various aspects that we can choose to monitor for behavioral changes. We focus on the following *four* criteria in choosing aspects to measure:

Objective Quantification We can quantify what we are measuring. For instance, determining a subject's workout quality using the the scale *intense*, *good* and *normal* is not very objective since the subject's perception might change during the course of the study especially as her stamina builds up. Instead, counting the number of steps she walks each day is more objective and quantifiable. This number also strongly reflects the workout quality.

Easily measurable We can measure what we are interested in without the need for expensive or bulky equipment. Some aspects might also require periodical measurements throughout the day. Thus, any aspect that we measure should not take up too much time or effort lest the subject is tempted to cheat or fabricate information. Whenever possible, an implicit health monitoring equipment would measure the aspect and minimize the work the subject has to do.

Non-intrusive Any measurement that inadvertently causes the subject to change her normal behavior should be avoided. For instance, using a electrocardiogram allows a subject to easily determine her heart rate. However, while being accurate and precise, an electrocardiogram is an intrusive device. The subject needs to be hooked up to such a device which hinders her normal daily activities, causing her to deviate from her normal behavior.

Universal Any aspect that we choose should be applicable to as many subjects as possible. For instance, while it is easy to quantify how many hours a subject runs on a treadmill in a non-intrusive manner, that aspect is only applicable to subjects who workout on a treadmill; many subject do not have this habit.

By selecting aspects that are *easily measurable* and *non-intrusive*, we aim to minimize the effects of the Hawthorne Effect [10] where the subject behaves *differently* because she realizes that she is under observation.

2.1 What To Measure

Based on our criteria above, we have selected the following three *universal* categories for monitoring and measurement. For each category, we describe the aspects that we are measuring and the rationale behind our choice.

Diet We measure the times when subjects eat their meals. We conjecture that a non-stressed person tends to eat her meals regularly without skipping any. A stressed person, on the other hand, has less time to sit down to eat a proper meal and would thus have irregular eating times and would eat when *possible* i.e. when there is a break in her schedule [4]. Some studies suggest that chronic stress might also cause one to overeat and consume more high-fat foods [1]. In short, we believe stress has an observable effect on a subject's eating habits.

To determine the quality of a subject's meal, we measure the different colors in it. We base our hypothesis on the Japanese idea of *go-shiki* [2], where a healthy meal usually consists of five different colors. We conjecture that a stressed person pays less attention to her food and would thus opt for a quick meal such as an energy bar without many different colors.

Exercise We measure the number of steps that subjects take each day. To determine the quality of steps, we measure the number of aerobic steps (see Section 2.2). We conjecture that a stressed person has different walking habits compared to when she is not stressed. For instance, an office worker might walk less when she is stressed because she is busy working in front of the computer. On the other hand, she might instead move around more when she is stressed if she has a lot of errands to run. *Either* scenario is possible, yet both demonstrate a deviation from the normal number of steps during non-stressful times.

Sleep We measure the times when subjects sleep and wake up. With this we can calculate the number of hours that they sleep for the night. We conjecture that a stressed person tends to get insufficient sleep each night. A 2002 poll conducted by Gallup [11] and a 2009 poll [14] by the American Psychological Association corroborate this conjecture by reporting that many Americans were losing sleep because of stress.

To determine the *quality* of the sleep, we ask subjects to record if they felt their sleep was *restful*. A restful sleep implies that a subject wakes up feeling well-rested. A unrestful sleep implies that the subject wakes up feeling that she needs more sleep.

We also ask subjects to record two additional *fuzzy* attributes about their sleep: if they recall dreaming in their sleep **and** if they recall being interrupted in their sleep. These attributes are *fuzzy* because they do not fit into our *Objective Quantification* criterion discussed in Section 2. Nonetheless, we believe that these two attributes might yield more insight into the behavioral patterns of our subjects.

For dreams, subjects record the *nature* of their dreams as either *good*, *neutral* or *bad* (or *unsure* if they were uncertain about whether or not they dreamed during a sleep period). We are interested in understanding any connection between the nature of subjects' dreams and stress [1].

For interruptions, subjects record if they recalled being interrupted during their sleep. Recording interruptions is a useful *check* to determine if the subject's unrestful sleep was due to actual stress or due to uncontrollable external interruptions. For instance, a subject might live in a very noisy environment and constantly be interrupted throughout the night; her unrestful sleep might not be from stress but because of her environment. By recording this information down, we could use it while interviewing the subject to verify why she was not sleeping restfully.

We monitor multiple aspects for each category because some aspects are stronger indicators of stress in a particular person. For instance, Sam's eating habits might change while he is stressed but his sleeping habits might not. Therefore, by measuring multiple aspects we hope to find at least one that correlates better to stressful behavior. By experimenting with less conventional aspects such as the colors in food and types of dreams, we hope to uncover interesting aspects that have never been discovered before.

[1] It is important to note that we are **not** interested in the *interpretation* of dreams.

2.2 How To Measure

Deciding how to measure is limited by the current devices that we have and also how intrusive that device is. We prefer non-intrusive devices whenever possible. Non-intrusive devices minimize the behavioral changes in a person and allow us to collect data that is more truthful and representative of normal behavior.

We present our choice of technology for each category below and the rationale behind our choices.

Diet Our subjects will use their cellphones to take a top down picture of their meals. Cellphones are ubiquitous devices that almost everyone carries; thus we can expect that most subjects will already have one and will not consider its usage to be intrusive [17].

Cellphones are equipped with built-in cameras that can take acceptable photos. Each photo also has a timestamp associated with it, which we use to determine the time of the meal. To determine the different colors in a meal, we perform basic image processing on the photos.

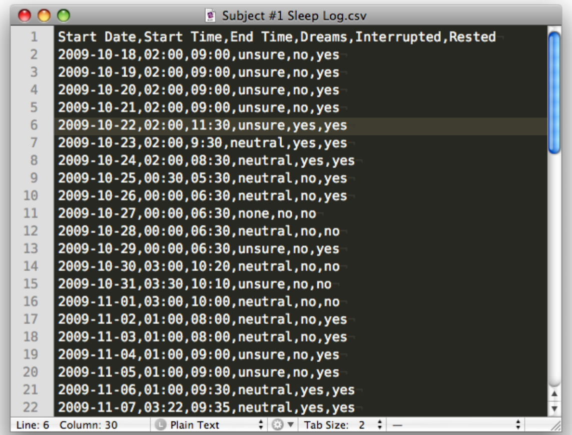
Exercise Our subjects will carry the Omron HJ-720ITC pedometer (see Figure 1) everyday. It is small, light and fits easily into a pocket or a purse while still measuring the number of steps taken reliably; it is a non-intrusive device. This particular pedometer records the number of steps for each hour. In addition, it also records *aerobic steps*. Aerobic steps are recorded when the subject takes more than 60 steps per minute and walks continuously for more than 10 minutes.

Figure 1: Omron HJ-720ITC Pedometer.

Sleep Our subjects will record their sleeping times in a log file. A sample log is shown in Figure 2. While devices exist to record certain attributes of sleep, we feel that they are too bulky and expensive to require our subjects to use. Keeping a sleep log requires some inconvenience on the part of the subject, but we feel that this was justified as most subjects only sleep once a day and thus only record a single entry each day.

Contrast this to using a specialized device. While a specialized device can give more fine-grain information, it might not be portable. Thus, we cannot collect data when subjects travel and sleep at different places. Not having continuous data poses a significant problem when trying to find changes in behavior.

Moreover, currently no devices exist that are capable of determining whether a dream is *good*, *neutral* or *bad*. We still need to rely on our subjects to provide that information.


Subject #1 Sleep Log.csv

```
Start Date,Start Time,End Time,Dreams,Interrupted,Rested
2009-10-18,02:00,09:00,unsure,no,yes
2009-10-19,02:00,09:00,unsure,no,yes
2009-10-20,02:00,09:00,unsure,no,yes
2009-10-21,02:00,09:00,unsure,no,yes
2009-10-22,02:00,11:30,unsure,yes,yes
2009-10-23,02:00,9:30,neutral,yes,yes
2009-10-24,02:00,08:30,neutral,yes,yes
2009-10-25,00:30,05:30,neutral,no,yes
2009-10-26,00:00,06:30,neutral,no,yes
2009-10-27,00:00,06:30,none,no,no
2009-10-28,00:00,06:30,neutral,no,no
2009-10-29,00:00,06:30,unsure,no,yes
2009-10-30,03:00,10:20,neutral,no,no
2009-10-31,03:30,10:10,unsure,no,no
2009-11-01,03:00,10:00,neutral,no,no
2009-11-02,01:00,08:00,neutral,no,yes
2009-11-03,01:00,08:00,neutral,no,yes
2009-11-04,01:00,09:00,unsure,no,yes
2009-11-05,01:00,09:00,unsure,no,yes
2009-11-06,01:00,09:30,neutral,yes,yes
2009-11-07,03:22,09:35,neutral,yes,yes
```

Line: 6 Column: 30 Plain Text Tab Size: 2

Figure 2: Sleep log recorded in a text file showing times of sleep and restfulness of sleep.

In the ideal scenario, all these measurements would be done using automatic health monitors which require little or no intervention from a subject. However, technology is not quite there yet without being prohibitively expensive. The number of devices required has been kept to a minimum. Everyday, a subject only needs to remember to carry a cellphone, which she normally carries, and a pedometer. She also needs to remember to record the times when she sleeps and wakes up.

We feel this is an acceptable level of intrusion on a subject's daily life.

3. EXPERIMENT DESIGN

In our current study, we acted as the subjects, monitoring ourselves for a period of seven weeks. We measured our stress levels for those seven weeks and attempted to form some correlation between our stress levels and our observable behavior for diet, exercise and sleep. From this point forward in the report, we refer to ourselves in the third-person when referenced as the subjects of our study.

3.1 The Subjects

Subject #1 A 42 year-old American, African-American ethnicity, male graduate student in Computer Science, living in a home with his wife and toddler.

Subject #2 A 26 year-old single New Zealander, Korean ethnicity, female graduate student in Computer Science, living alone in off-campus housing.

Subject #3 A 26 year-old single Malaysian, Chinese ethnicity, male graduate student in Computer Science, living in off-campus housing with a housemate.

3.2 The Duration

Table 1 shows the timeline of our study and notes any significant events that might contribute to increased or decreased stress levels. For instance, Subject #1 has a two-year old toddler who has not adapted to sleeping through the night; thus he has to wake up frequently in the middle of the night to look after her. Subject #2 had a two-week long vacation where she went home to visit her family. Subject #3 is not fond of traveling and feels stressed when he has to travel for a conference.

3.3 Measuring Stress

To determine the stress levels of our subjects for the seven weeks, we had them take the NSAD Stress Questionnaire by the International Stress Management Association UK [7]. Some of the questions are shown in Table 2 as reference. The number of "yes" and "no" answers are tallied and compared against the provided stress scale.

Questions	Yes	No
I frequently bring work home at night		
Not enough hours in the day to do all the things that I must do		
I deny or ignore problems in the hope that they go away		
I do the jobs myself to ensure they are done properly		
I underestimate how long it takes to do things		
. . .		

Table 2: Sample Questions from the Stress Questionnaire.

Taking the questionnaire was the simplest way to measure a subject's stress level. However, this method was not without problems. We only administer the questionnaire once a week because we were worried about desensitizing the subjects to the questions; if we had them take the questions everyday the subjects will become too accustomed to the questions and probably not answer them carefully.

Because we only administer the questionnaire once a week, we could only use the week as our period of measurement. Stress is a very volatile emotional state: it might be as fleeting as a few minutes or it might continue for several days. Nonetheless, we had no way to tell because our questionnaire was only administered once a week.

To account for this, we interview the subjects to record any significant events that might have occurred during the week and determine if those events could be possible causes of stress.

Week #	Dates	Significant Events
Week 1	Oct 18 - Oct 24	
Week 2	Oct 25 - Oct 31	Subject #3 attended OOPSLA Conference 2009
Week 3	Nov 1 - Nov 7	Subject #1's child starts sleeping well through the night Daylight Savings Time starts
Week 4	Nov 8 - Nov 14	
Week 5	Nov 15 - Nov 21	Subject #2 on vacation overseas
Week 6	Nov 22 - Nov 28	Subject #2 on vacation overseas Subject #1's child's sleep schedule thrown off Thanksgiving Break
Week 7	Nov 29 - Dec 6	This project report was due

Table 1: Duration of the Seven Week Study.

Week #	Subject 1	Subject 2	Subject 3
Week 1	Stressed	Not Stressed	Not Stressed
Week 2	Stressed	Stressed	Stressed
Week 3	Stressed	Stressed	Not Stressed
Week 4	Stressed	Stressed	Stressed
Week 5	Stressed	Stressed	Stressed
Week 6	Stressed	Not Stressed	Not Stressed
Week 7	Stressed	Stressed	Stressed

Table 3: Weekly survey results for each subject.

4. RESULTS

Although this study was carried out only on a population of size three over seven weeks, our results show some very interesting and clear connection between daily behaviors - diet, exercise and sleep - and the stress levels. Note that we have not performed any form of formal analysis on the data collected but we plan to employ statistical analysis in our future work as we will increase the population size considerably. Section 4.1 shows some noteworthy results per subject. We consciously chose not to compare the results between different subjects because individuals react to stress in different ways and it is infeasible to define a norm in eating, sleep or exercise patterns. The data collected were analyzed based on the individual's weekly survey results that was used to indicate the stress level (Table 3).

Note that the horizontal axis in each graph represents a 24-hour time frame starting at midnight.

4.1 Individual Results

Subject #1 was a special case as he scored high on all of the weekly stress surveys and thus marking every week as stressful. This was an unexpected but entirely possible scenario. As a result, we could not draw a reliable correlation between behaviors and the level of stress for Subject #1 because we were not able to determine what his normal behaviors are during unstressful weeks.

Subject #2 's most prominent indicator of stress was observed in exercise data, specifically the in the patterns. The total number of steps per week varied only little with no particular trend, but during the weeks marked stressful, the most active period of exercise occurred very regularly early in the day. This was when Subject #2 walked to work and it was confirmed after interviewing Subject #2, that she tends to have a strict commute schedule when under heavy work load (Figure 3). A notable pattern was also found in sleep data, where Subject #2 had considerably longer sleep segments in unstressful week compared to stressful weeks. The remaining criteria, diet, did not show any obvious patterns. This was because Subject #2 has irregular eating habits and usually has less than three meals a day, so skipping a meal or having meals at unconventional time was not an indicator of stress. This again was confirmed after an interview with the subject.

Subject #3 had the strongest stress indication in his sleep data. He had relatively more nights of unrestful sleeps, up to four nights which are represented in purple bars, during stressful weeks compared to only up to one night in unstressful weeks (Figure 4). Subject #3's sleep pattern did not changed much over the course of study whether he was stressed or not, but by keeping a record of quality of sleep, we were able to deduce this correlation. Subject #3's diet or exercise data showed less obvious patterns. The fact that he lives very close to where he works, and also that his work usually has him sitting down at a computer, may be some of the factors that lead to no clear connection between his exercise and stress levels. As for diet, even though there were few outliers during his stressful weeks, the separation was not clear enough to decide that they were caused by stress.

4.2 Remarks

There were more data collected than we present in this report. For example, we initially recorded more characteristics of sleep, such as whether a subject had (good or bad) dreams, and if sleep was interrupted. This data however was not included in our analysis as we were not sure how to intergrate it. For exercise data, we recorded of the number of aerobic steps as well as the total number of steps, but the pedometer's definition of aerobic steps - 10 minutes of continuous walking with more than 60 steps per minute - was too limiting and we felt that it did not convey the quality of walk very fairly. Lastly, the result we obtained from histograms of food colors showed little variation between stressful and unstressful weeks, with a large proportion of food being orange, brown or tan hues.

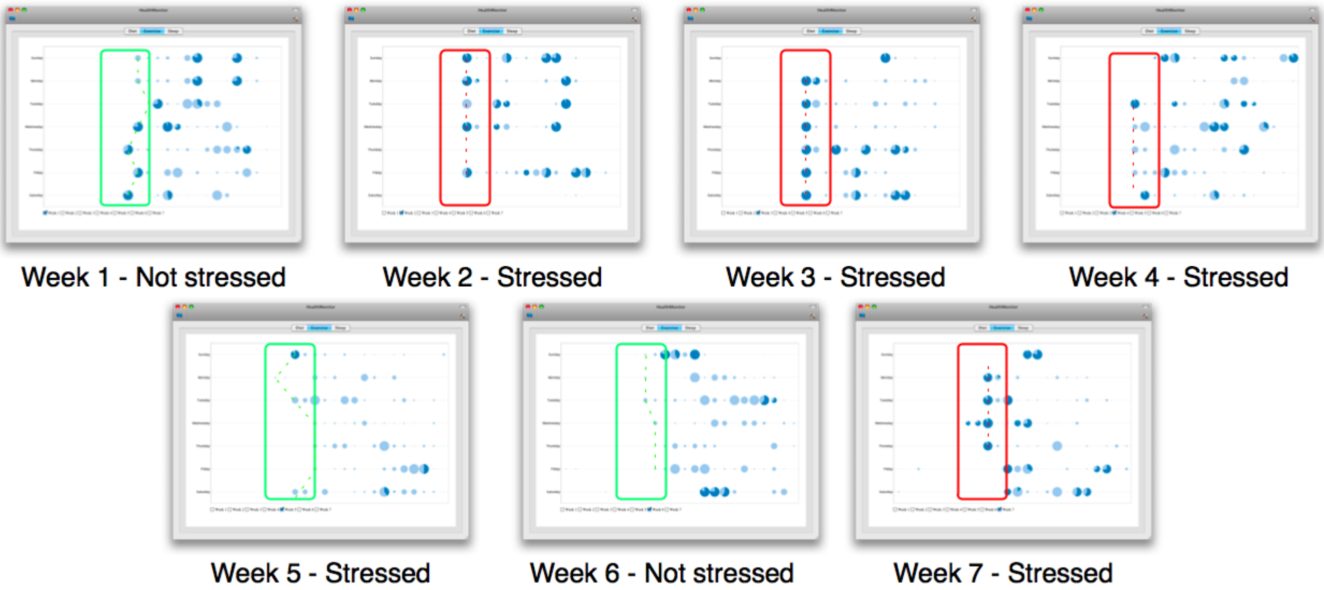


Figure 3: Subject #2's Exercise Data - size of each dot represents number of steps, and darker wedges represent aerobic steps.

Figure 4: Subject #3's Sleep Data - green bars show restful sleep, purple bars show unrestful sleep. Stressful weeks have comparatively more number of unrestful sleep than unstressful weeks.

5. OBSERVATIONS

Since the subjects were also the designers of the experiment, they were highly motivated to be "good" subjects. They were meticulous about wearing their pedometers as much as possible, taking photos of their meals, and logging their sleep periods. Participating in the study was far more challenging for subjects than we imagined. For a variety of reasons, each of the subjects failed, on multiple occasions, to properly collect the required data. Our experience as subjects was invaluable in understanding first-hand the challenges of being a participant. We gained insight into better designing an experiment in which everyday people are expected to perform tasks on a regular, super-daily basis. To some degree, all of the subjects succumbed to the Hawthorne Effect [10], and more generally to the phenomenon of *reactivity* [5] - also known as the Observer Effect [16]. Finally, the subjects experienced emotional reactions to being participants in the study that we had not anticipated. Detailed below are the observed challenges for each measured aspect.

5.1 Sleep Challenges

Accurate logging of sleep patterns was compromised by issues of fatigue, instrumentation, and subjectivity.

5.1.1 Fatigue

It was not uncommon for the subjects to be very sleepy when going to sleep or awaking. When subjects were sleepy when falling asleep, it was easy for them to forget to record the time, or simply pass out before they were able to record the commencement of sleep. Likewise, when awaking, particularly when sleepy or awakened with a start, it was easy for them forget to record the event. Conversely, there were occasions when subjects attempted to go to sleep, but did not actually fall asleep at their initially recorded time. Finally, regardless of whether the subject successfully recorded the commencement of sleep, it was all but impossible for them to be certain the that the record time was the actual time. Without instrumentation, is impossible for persons to know *exactly* when they fall asleep. It is always an estimate at best.

5.1.2 Instrumentation

In order to record the beginning and ending of sleep periods, the subjects had to have an accurate time source nearby. Sometimes this was not possible. For example, Subject #2 awoke from a nap while flying across the Pacific. Even with a watch handy, considering the time zone and international timeline, she had no idea what time to record. The subjects also needed to have an implement ready to record the times; e.g. an iPhone, pen and paper, etc. If they didn't, they would have to remember to record the time later. Again, given such a memory had to be made when the subjects were tired, made such measurements inherently less reliable.

5.1.3 Subjectivity

Subjectivity was a further challenge. Upon waking, the subjects were to record the quality of their sleep by commenting whether their sleep was restful. Measuring the quality of a dream is highly subjective. Even if one is regularly aware of one's dreams, it is challenging for one to characterize the *nature* of one's dreams. Furthermore, even when one's dreams are vividly remembered, it is challenging for one to judge whether the dreams are good or bad. For example, Subject #1 woke up crying, recalling a dream about his deceased father. For many, this would easily be deemed a bad dream. However the subject marked the dream as good, because he misses his father and was happy to have an explicit dream about him, since he had not had one about him in sometime.

5.1.4 Underspecification

Though we did our best to define what an interruption of sleep was, as the experiment progressed it became obvious that our initial specification for interruptions was inadequate. We had considered events such as waking to go to the bathroom in the middle of the night, or awaking because of a crying baby; yet there was the ongoing ambiguity about what the maximum duration of an interruption before an interrupted sleep period should be considered two separate sleep periods. Also, we failed to provide the subjects clear guidelines about whether an expected interruption which terminated a sleep period (i.e. an alarm clock) should be considered an interruption or not.

Also, our central question, "do you feel rested", was not as straightforward as it initially seemed. For example, as Subject #1 was chronically fatigued, his notion of restfulness took on an ambiguous meaning. He reported that his measurement of restfulness was compared to other recent sleep, as opposed to compared to his normal or lifetime baseline of "good sleep".

Finally, though our visualization shows restful versus non-restful sleep, it currently fails to show the relationship of dreams and interruptions, to restfulness.

5.2 Diet Challenges

Challenges in recording meals largely fell into three categories: unintentional, behavioral, and instrumental. Of the three personal measurements, we knew in advance measuring diet would be the most demanding of task the subjects would be required to do. The study confirmed this to be the case. The subjects reported that it was far more challenging than any of us expected for both expected and unexpected reasons. In designing the experiment, we were aware of the obvious problem of the subjects' needing to have their camera-enabled mobile-phone at the ready, all the time. If a subject forgot their recording device, the record of their meal would be lost.

5.2.1 Unintentional

Unintentional failures to record meals included such event as: a subject forgetting to take a picture of their meal; not having their cellphone available to take a photo; or taking a picture of something the subject didn't (fully) eat.

5.2.2 Under-reporting

In practice, subjects reported few occasions where they missed reporting a meal because they didn't have their cameras. In fact, most under-reportings of meals occurred because a subject simply forgot to take a photo before finishing their meal. Forgetting was by far most likely to happen when the meal took place with others. Subject #1 reported that when he ate meals alone, he rarely forgot to take a picture of a meal. Conversely, when he was eating with others, it was an absolute struggle to remember to take a picture. In fact when eating meals with others, Subject #1 said that, more often than not, he got so engrossed in conversation that he completely forgot to take a photo, remembering only after meeting the mocking gaze of his empty plate. During the Thanksgiving break, while visiting family, several subjects lamented that they hardly recorded a meal at all.

Sometimes the missed meal could be represented by taking a photo of the same food items; e.g. if the subject ate an apple and cheese and forgot to take the photo, if they had those food items readily available, they could take a picture of the replacement items instead. The subjects repeatedly did this during the experiment. One downside of this is behavior was that the timestamp for the replacement photos was be later than the actual meals. Since the subjects frequently ate the same meals, another option for them was to find an earlier photo of the same meal, make a copy, and update its timestamp. Obviously, this approach is more work than the former approach. In practice however, none of the subjects reported using this approach. One subject reported that he *considered* using this technique, but would then forget and would let too much time pass after the meal to perform the repair reliably.

Another common problem was under-reporting of serial snacking. While subjects were only instructed to take photos of meals, frequently they had to make a nontrivial judgement call about when was their snack big enough to actually be considered a meal. For example, a handful of cereal out of the box is arguably a not a meal. However, several trips to one's cereal box over an afternoon could easily add up to more food that a single serving of cereal for one's breakfast. These *composite* meals are not accounted for in any regular way, but nonetheless these incidents happened from time to time.

Finally, looking through the record of photos, it is apparent the beverages were not included in the recording of meals in any regular or substantial way.

5.2.3 Over-reporting

Not infrequently, subjects would capture food that they didn't eat. This would typically happen when a subject made a plate of food but didn't actually eat all of items, e.g. a meal photo might improperly credit the *red* in a salad, even though the subject didn't actually eat any of the red pepper in the salad. Sometimes, the subjects reported eating only part of a meal because they were interrupted and never got to finish, or only finished the remaining items much later. This problem created an awkward dilemma for the subjects about whether when they ate the second half of a meal later, the meal should still be credited with the initial photo. If the subject decided to take a later photo of the remaining items when they actually ate them, it was ambiguous about what they should do about the initial photo which included the the uneaten items. This problem touches on another issue in the design of the study: what is the length of a meal, and does the length of a meal matter? The timestamp of the photos only captures an instantaneous moment, however a meal always occurs over a period time.

5.2.4 Consistency

Since the subjects were also the experimenters, they knew how to take photos in a consistent fashion that make analysis of the images easier. Nonetheless, the constantly changing eating locations, and background and lighting conditions made this challenging. Though the subjects were always able to shoot from the same position and angle, they were frequently unsure about how to crop the shots of the food. As a scaling index, the subjects were to include a penny in each photo. However, as common as pennies are, it was surprising how onerous this requirement was for the subjects. Not long into the study, we abandoned this requirement.

Another problem was food wrappers. Many pre-prepared food items have wrappers that needed to be *fully* unwrapped to display the food, which then unfortunately, often makes the food difficult to eat. Also, colorful wrappers were easily confused as food during the image analysis. Finally, with foods such as sandwiches or burritos, it was ambiguous to the subjects whether or not they should split these items open before photographing them, least the colorful ingredients be completely obscured by their tan and brown coverings.

5.2.5 Influenced Behavior

We imagined that instrumenting the subjects' meals via photography would, of the three measures, be most likely to influence the subjects' normal behavior. Our suspicions were indeed born out by the study. For example, subjects reported either: sometimes not eating when they were inclined to because they didn't have a camera ready; or not eating because the hassle of taking the photo felt like more trouble than the inconvenience of being a little hungry; or conversely the guilt of being a "bad" subject by eating without taking a photo was too high. On other occasions, subjects ate and *deliberately* didn't take photos because it would have been too awkward in their current social setting. Most unexpectedly, Subject #1 was frequently inhibited in his eating and photo-taking by the fact that his toddler was obsessed with his cellphone and camera. When in her presence, Subject #1 would have to surreptitiously photograph his meals.

Either out of curiosity or in order to "repair" missing meals (as discussed earlier), some subjects reviewed the photos of their meals. Being present to their eating in this explicit fashion undoubtedly had a conscious if not unconscious effect on these subjects further eating behavior. Suspecting this, at least one subject deliberately chose not to look at the photos at all, which created the unfortunate consequence for him of not having access to any other photos he had taken with his phone during the study. In any case, we underspecified whether or not the subjects were expected to review the photos of their meals.

5.3 Exercise Challenges

Of the three measures, the pedometer simultaneously represents the best and worst means of instrumentation. It was the best because, far and way, it was the least intrusive for the subjects. Conversely, since it was *so* invisible to the subjects, in was easy to be *completely* unaware that the pedometer wasn't attached at any particular moment.

5.3.1 Uncredited Exercise

If the subjects forgot to wear their pedometers, their data was lost completely. This was surprisingly easy for them to do when changing their clothes. Unlike with sleep and diet, there were no reliable ways for the subjects to fill-in missing data. Also, though the pedometers had a seven day memory, it was easy for the subjects to forget to off-load their data before it expired. By and large, when data was lost it was impossible to replicate.

Due the design of our chosen pedometers, it is likely that many of the subjects' aerobic steps were under-reported. Early on, subjects reported concerns that the pedometers were not crediting their aerobic steps. In particular, Subject #1 reported that he had been out walking all day one Saturday with his daughter but, surprisingly, had *zero* aerobic minutes. Learning this, we performed a test and discovered that the pedometers were indeed working properly, but that they only reported aerobic minutes when their wearers walked continuously for 10 or more minutes at a time. While Subject #1 had walked for many hours, he had only walked in periods of less than 10 minutes at a time.

The pedometers also frequently missed steps while the subjects were at home, because, as a rule, the subjects only wore the devices with their street clothes. In particular, the pedometers missed valuable steps in early morning and late evening, or frequently when the subjects were working from home.

5.3.2 Influenced Behavior

Though the pedometers were for the most part, profoundly non-intrusive, occasionally they frustrated their wearers. Subjects reported feeling angry that the devices had *cheated* them of their "good" aerobic steps. Subjects were frustrated when for example, they ran a quick errand home, while in their pajamas but without the pedometer, and "lost" steps. They reported being tempted to shake the pedometer in order to "regain" lost credit.

As with the photos of their meals, the subjects could be influenced by reviewing their walking logs from their pedometers. Again, as with the photos, some subjects reviewed their logs while others did not.

Finally, when we ran our test on the pedometers, discovering about the devices' 10 minute requirement for aerobic exercise, we introduced *unnatural* behavior into the subjects' data. This is notable because Subject #1 had only *two* recorded aerobic sessions during the entire study; our test being one of them.

5.3.3 Consistency

Our choice of exercise instrumentation caused one notable major inconsistence issue. Our study spanned the ending of daylight saving time and it didn't immediately occur to us that the pedometers didn't automatically adjust to the change in time. This oversight required us to perform some annoying post processing on some of the data, until the pedometers internal clock were readjusted.

6. FUTURE WORK

This study was an invaluable opportunity for us to experience firsthand the personalization that is often missing in population studies. Our preliminary results and observations provide a strong basis for the design of our next study.

6.1 Instrumentation

We will make a number of changes to the devices the subjects use to record their diet, exercise and sleep data. We will also continue to employ a number of manual instrumentation methods.

6.1.1 Meal Photography

As discussed, photographing their meals was the most labor intensive aspect of the study for the subjects. Also many subjects missed recording meals every now and then. In the next study, we will require the subjects to use smartphones with built-in cameras, and will install our custom software on their phones that assist with their photo taking, management and repair process. We will write a software application that, immediately after a photo is taken, would enable the subject to crop, scale, white balance, tag, and otherwise prepare the images in a more consistent fashion. When a subject realizes that they have missed taking a photo of a meal, the application will assist them in preparing a replacement from an existing or new image.

Also, in standardizing on the use of smartphones, all images will be GPS tagged, so that we will also have new data about *where* the subjects' meals occur.

6.1.2 Heart Monitors

Though the pedometers were very non-intrusive, we are open to using different measuring devices in concert with, or in replacement of the pedometers. In particular, we are considering switching to using portable heart monitors which are used to measure heart rate. While heart monitors are usually more expensive than pedometers, they have a number of important advantages.

They connect to a subject on their wrist and/or via on a band around their chest, and thus are **not** dependent upon clothing. This will enable us to capture exercise during times when the subjects are typically under-clothed; times that would be under-reported using pedometers. Since heart monitors measure the heart rate *directly*, they can more easily be connected to subjects' aerobic activity, and will implicitly have finer-grain distinction of heart rate than the pedometers.

Additionally, we may be able to "kill two birds with one stone" in that we should be able to use the heart monitors to instrument sleep as well. Human beings have distinctly lower heart rates when they are at rest, and lower still when asleep.

6.1.3 Pedometers

Under any circumstances, if we do continue to use pedometers, we are strongly considering replacing our current model with another model or brand that also allows easy computer connectivity, but doesn't use an arbitrary cut-off time (e.g. 10 minutes) to distinguish aerobic and non-aerobic steps.

6.1.4 REM Monitors

The dream data from our preliminary study was too subjective in correlating the quality of dreams to stress. In order to record dream activity more objectively, for the next study, we are considering using piezo-electric film attached to the subjects' eyelids while they sleep at night to detect and record REM activity. Our concern about this approach is that this form of instrumentation may be too intrusive for a longterm study.

6.1.5 Sleep Logs

In spite of our concerns about the subjectivity, in the next study we still plan for subjects to record their quality of sleep and nature of dreams data; along with other supporting objective data, these may yet prove to be a fruitful measures. To assist subjects, upon waking, they will use our custom appliation on their smartphones to record they their sleep and dream information. A short menus will enable them to quickly enter the period of sleep, and their evaluation of their sleep and dreams. Comparing the subjects' estimations of their sleep periods and presence of dreaming, against that provided by the instrumentation (i.e. heart and REM monitors) may provide some new insights.

6.1.6 Stress Tests

Our initial approach to measuring stress was inadequate. One test per week was too large a granularity to be useful. Using the subjects' smartphones, we will deliver the stress questionnaire more frequently. The questions will be more nuanced so that the responses are less likely to be fully saturated as *stressed*. This should enable us to gather more detailed data which may show a clearer correlation between the subjects' daily behaviors and stress levels.

6.2 Management of Subjects

We will use careful consideration in selecting the next batch of subjects for the study. We will give them more explicit instructions and debrief them at the end of the study.

6.2.1 Selection

We will seek to recruit subjects that live and work near at least one other subject. As subjects, we benefited tremendously by having each other nearby to support and encourage each other in taking photos and maintaining our logs. Without the support of other participants, family and coworkers, subjects are far less likely to participate fully.

6.2.2 Instructions

We now have a better idea of how to provide good instructions to the subjects. Our goal is for them to have fewer judgement calls and less anxiety about measuring their behaviors. We will provide clearer instructions to the subjects; in particular, about staging photographs of food, and clearer guidelines about differentiating meals from snacks. We will also provide them with access to a detailed FAQ of common issues.

Given the choice of having subjects be aware or stay unaware of their data, we will be explicit in *encouraging* them **to monitor their own data frequently**. Though this approach may influence the subjects' behavior, the effect of this influence may be measurable and pertinent. Importantly, it will make it more likely that the subjects have less holes in their data.

6.2.3 Debriefing

After the study, we will conduct exit interview with the subjects to benefit from their *own* personal observations; much as we did in our preliminary study.

7. RELATED WORK

There have been various efforts in designing personalized implicit health measurements. All of these, however, require that the subjects reside in specially designed environments that continuously collect behavioral data. Once the subjects have moved into these environments, the monitoring process is fairly non-intrusive. Nonetheless, the act of relocating to this new environments might be considered the *most* intrusive requirement since the subjects have to move into those specialized homes to be monitored.

The *Aware Home Research Initiative* at the Georgia Institute of Technology [9] attempts to measure all aspects of behaviors of occupants living in a specially instrumented home. This is a long-term project that attempts to study and document the behaviors of people in the hopes of creating a smart home that can adapt to its habitants needs. Our project, on the other hand, tries to monitor and measure behavior using simple non-intrusive devices. Our approach is more general and scalable because it works without requiring subjects to live in a special building. However, because our approach does not constantly monitor all aspects of the subjects' daily lives, we may fail to detect subtle, but nonetheless important, changes in behavior.

The *Ambient Intelligence* project by Philips [15] attempts to create a smart home with high-tech devices that will monitor all aspects of its occupants behavior. Its main goal is to create a smart home that will adapt to its occupants needs and moods and adjusting to it. Like the *Aware Home Research Initiative*, this project requires that its subjects live in a specially instrumented home. While this is certainly feasible in a few decades' time, the current technology is prohibitively expensive to deploy today. Correlating stress with behavior is conceivably a possible goal of the *Ambient Intelligence* project but currently the project is more focused on trying to create a smart home that will adapt to its users needs.

Intel has various projects on smart homes with various different goals including trying to correlate behavior with different illnesses. However, details of their projects are sparse. Some of their projects are mentioned on their website [6] and in Stephen Baker's book [3]. Without more information, we are unable to report on the similarities and differences between their various projects and ours.

There are also existing work on using non-intrusive devices to measure human behavior. Pentland at the MIT Media Lab leads various *Reality Mining* projects that utilize non-intrusive, wearable sensor devices; they then mine the collected data for various purposes including healthcare [13]. His group has had success with implicit sensors that monitor various aspects of human behavior which they then use to correlate and predict human behavior. His work is closely related to ours since it utilizes non-intrusive, personalized implicit monitors. His group has not tried relating human behavior with depression or stress.

8. CONCLUSION

A person's psychological well-being can be deduced by observing her past, current and future behaviors. Specifically, changes or patterns in her behaviors can be seen as an indication of changes in her psychological health. In this study, we introduce *personalized implicit health monitors* to observe correlations between a person's daily behaviors and her level of stress. We chose stress as our target health issue because most adults experience some form of stress from work, school or their personal lives on a daily basis. In addition, stress can be viewed as a milder form of anxiety, which in turn is a milder form of depression - a more serious health problem.

Our method is *personalized* – individuals vary in ways they react to stress. It is infeasible to define a norm in humans' behaviors and apply it to a population in order to classify a deviation as either typical or atypical. Our method is also *implicit* – subtle changes in behavior are implicit indicators of stress in a person. And because our method uses easily measurable aspects of behaviors – diet, exercise and sleep in non-intrusive manners – we minimize changing the subjects' typical behavior via instrumentation and have more truthful and representative data.

However, try as we might, there is no avoiding having some effect on monitored subjects. Our preliminary study provided key observations to the challenges of instrumenting real people. These observations provide meaningful insight to enabling a consistent and non-intrusive approach to further *personalized implicit* health monitors studies.

9. REFERENCES

[1] T. C. Adam and E. S. Epel. Stress, Eating and the Reward System. *Physiology and Behavior*, 91(4):449–458, July 2007.

[2] E. Andoh. *Washoku: Recipes from the Japanese Home Kitchen*. Ten Speed Press, 2005.

[3] S. Baker. *The Numerati*. Houghton Mifflin Harcourt, 2008.

[4] E. A. Charlesworth. *Stress Management: A Comprehensive Guide to Wellness*. Ballantine Books, 2004.

[5] P. P. Heppner, D. M. Kivlighan, and B. E. Wampold. *Research Design in Counseling*. Wadsworth Publishing, 1998.

[6] Intel Health. Intel Healthcare Research Portfolio. Website. http://www.intel.com/healthcare/research/portfolio.htm?iid=health+lhn_hriportfolio.

[7] International Stress Management Association UK. NSAD Stress Questionnaire. Website, 2009. http://isma.org.uk/pdf/nsad/NSAD-Stress-questionnaire.pdf.

[8] R. Kessler, W. Chiu, O. Demler, and E. Walters. Prevalence, Severity, and Comorbidity of Twelve-month DSM-IV Disorders in the National Comorbidity Survey Replication (NCS-R), 2005 June.

[9] C. D. Kidd, R. Orr, G. D. Abowd, C. G. Atkeson, I. A. Essa, B. MacIntyre, E. D. Mynatt, T. Starner, and W. Newstetter. The Aware Home: A Living Laboratory for Ubiquitous Computing Research. In *CoBuild '99: Proceedings of the Second International Workshop on Cooperative Buildings, Integrating Information, Organization, and Architecture*, pages 191–198, London, UK, 1999. Springer-Verlag.

[10] E. Mayo. *The Social Problems of an Industrial Civilization*. Routledge & Kegan Paul, 1975.

[11] D. W. Moore. Eyes Wide Open: Americans, Sleep and Stress. Website, February 2002. http://www.gallup.com/poll/5314/eyes-wide-open-americans-sleep-stress.aspx.

[12] National Institute of Mental Health. Depression. Website, 2008. http://www.nimh.nih.gov/health/publications/depression/nimhdepression.pdf.

[13] D. O. Oldguin, P. A. Gloor, and A. Pentland. Wearable Sensors for Pervasive Healthcare Management. In *3rd International Conference on Pervasive Computing Technologies for Healthcare*, 2009.

[14] M. Park. Why We're Sleeping Less. Website. http://www.cnn.com/2009/HEALTH/03/04/sleep.stress.economy/.

[15] Philips Research. Ambient Intelligence. Website. http://www.research.philips.com/technologies/projects/ami/background.html.

[16] L. Sechrest and R. J. Hill. Unobtrusive measures. In *International Encyclopedia of the Social & Behavioral Sciences*. Pergamon, 2001.

[17] J. C. Smith. The Efficacy and Feasibility of Mobile Phone Based Management of Chronic Illness. Master's thesis, University of Illinois at Urbana-Champaign, 2009.

[18] WHO. Constitution of the World Health Organization- Basic Documents, Forty-fifth edition, Supplement. Website, 2006. http://www.who.int/governance/eb/who_constitution_en.pdf.

APPENDIX

A. VISUALIZATION TOOL

This appendix explains the visualization tool used in the study. The tool was built from scratch by the three authors. The tool shows a visual analysis of data pertaining to a single subject at a time, in accordance with the personalization goal of the study. The visualization techniques used by the tool were designed so that any pattern in behaivors stand out.

The complete set of results for each subjects is available in the project repository, hosted at http://bitbucket.org/vazexqi/healthmonitor/src/tip/Data/.

A.1 General Layout

Figure 5 shows the general layout of our visualization tool.

1. **Analysis Selection Tabs** The tool gives a user an option to select which analysis to visualize. The tabs are present in all visualization windows, and each window

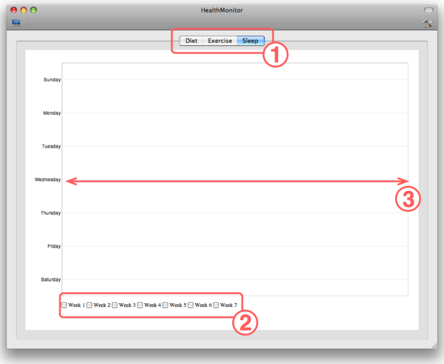


Figure 5: General layout of our visualization tool.

performs different analysis on an input file or folder selected. For diet data, user is prompted to select a folder containing all the pictures.

2. **Week Selection** A user can select which week's data to visualize. If multiple weeks are selected, the graphs are overlaid to show combined data points. This functionality is common across all visualization targets.
3. **Graph Axis** The horizontal axis of graphs (for diet, exercise and sleep) is a 24-hour time frame, starting from midnight. Using this representation it is easy to see any patterns in a subject's diet, sleep or exercise behaviors.

A.2 Diet

Diet graph analyses the pictures of meals taken by a subject (Figure 6).

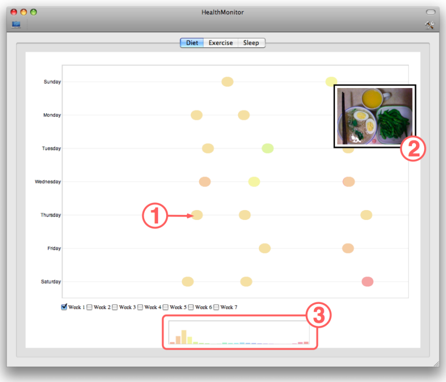


Figure 6: Screen capture of diet visualization screen.

1. **Diet Data** Each dot in a diet graph shows the day and time that the specific meal was consumed, and it also shows the most prominent color in the picture taken.
2. **Thumbnail** Hovering a mouse over each dot in a diet graph shows a thumbnail of the original picture of a meal that the dot is representing.
3. **Histogram** The histogram for diet data shows colors that are present in pictures. If multiple weeks are selected, the histogram shows a combined data for the pictures from those weeks.

Figure 7: Screen capture of exercise visualization screen.

A.3 Exercise

Exercise graph analyses the number of steps a subject has taken in hourly granularity (Figure 7). The pedometer used in the study records number of steps taken per hour, as well as the number of aerobic steps in that hour.

1. **Diameter of a dot** Each dot in the exercise graph represents the number of steps taken per hour. That is, the bigger the dot, the larger the number of steps taken. The diameter is saturated at 2000 steps, with an assumption that 2000 steps per hour is a reasonable estimation of maximum steps a person can take.
2. **Aerobic steps as darker wedges** Each dot is also a pie chart showing the percentage of aerobic steps over the total steps taken per hour.

A.4 Sleep

Sleep graph shows the sleep pattern of a subject (Figure 8).

1. **Restful and unrestful sleep** The color of each bar indicates whether the block of sleep was restful (green) or unrestful (purple). The length of each bar indicates the duration of sleep.

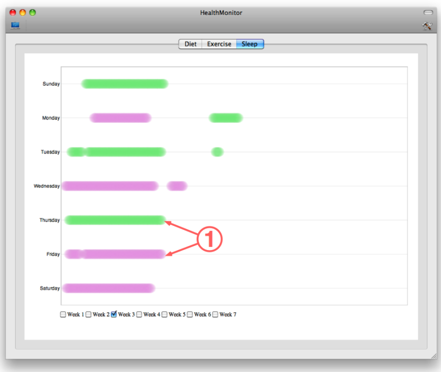


Figure 8: Screen capture of sleep visualization screen.