Maintaining Healthy Behavior Change Through Ubiquitous Sensing and Persuasive Feedback – A Pilot Study

Frank Chen, Jesse Cirimele, Maria Sandberg, Eric Hekler, Scott Klemmer, et al.

Stanford University

Stanford, CA 94305

fxchen@stanford.edu

ABSTRACT

Personal informatics through mobile phone sensing can play an important role in influencing healthy behavior. We study three systems in the field by prototyping upon current technologies with psychological methods. We study how different feedback and design technologies can influence an individual to maintain behavior change and to make behavior change actionable.

ACM Classification: H5.2 [Information interfaces]: User Interfaces, H1.2 [User-Machine Systems]: Software psychology

General terms: Design, Experimentation

Keywords: Guides, instructions, formatting

INTRODUCTION

Personal informatics and reflection has been an emerging trend in computing in the past years through ubiquitously connected mobile devices. Mobile sensing technologies have allowed sophisticated inferences through a variety of physical modalities. Research in the past has demonstrated a trend in using mobile computing for persuasive technology and helping people change their life styles [5]. A three month longitudinal twelve-person user study has shown behavior can be changed through mobile persuasion techniques [8]. Psychology and persuasion techniques show the best way to influence behaviors is by making the change maintainable and actionable [6]. Our research explores a mobile interface for instant feedback and reflective feedback, and goal setting feedback in order to improve behavior change technology.

Throughout this paper, we will start with what has motivated our research from the health science and persuasion technology fields. Then we will discuss our methodology in pilot testing two hypotheses through an Android widget application. Finally, drawing upon this study, we will discuss insights drawn from this pilot study to motivate future design and user studies to generalize our findings. The future study design is to observe longitudinally how large numbers of participants on the Android Marketplace interact with the system.

BACKGROUND

Health Guidelines & Health Sciences Group

Although sixty-five percent of Americans are overweight, studies show most Americans feel fine about their physical and eating habits [1].

Our project is a sub-study within the MILES (Mobility for Improving Life and the Environment at Stanford) group. The high-level goal of the group is to encourage people to lead more active lifestyles through mobile sensing and persuasion techniques.

In an earlier empirical study, we validated our capture method and accuracy with current best practice in the health sciences for capturing activity level through accelerometers and pedometers.

Mobile Sensing Research

Currently, the state of the art in personal activity level informatics is Ubifit [4], a mobile sensing and feedback application also aimed at healthy behavior change. This project had the insight to demonstrate reflection and motivation on an individual's physical activities through feedback on a mobile device home screen. This feedback took form in a garden widget on their cell phone home screen with indications of physical activities such as walking, biking, running, or stationary represented by different color flowers.

This allows an individual to look back upon their week and reflect upon their level of activity. This individual must decide for themselves what their goal for each week is. [7] implies goal setting was a difficult feature to use due to the cognitive load required.

METHODOLOGY

We aim to study walking as the primary measure of activity. To study the parameters to vary for measuring activity, we deployed three widget-based applications to test our two hypotheses, using Ubifit as our baseline. These three applications are visualizations we asked individuals to install on Android phone home screens. The home screen is a display an individual with this phone will view every time the screen turns on.

Our study design has two principle hypotheses:

> H1: Explicit goals for an individual's weekly activity and just-in-time feedback can increase the

engagement with a persuasive technology and increase walking time.

Because goal setting is an inherently hard problem for a individual to manage, we designed with the National Health guidelines in mind. Instead of allowing the individual to choose what the goal for the week, our design imperative becomes to state an explicit goal for the individual to walk towards.

The Ubifit garden metaphor relies upon reflection as a means to feedback activity level. Direct Manipulation research implies a feedback on activity level could bridge the gulf of execution and gulf of evaluation on physical activity. Our design imperative becomes to feedback the level of activity to the user.

> H2: Scheduling of activity times increases self-efficacy and the motivation to do physical activity, thus increasing walking time.

Psychology literature has proven self-efficacy can confirm beliefs in accomplishing activities an individual sets for themselves. Our design imperative becomes to build a scheduling system for the application to remind the individual when to exercise.

User Study
Through the course of seven days, we deployed three applications on one graduate students and a post-doc. We sorted individuals randomly into four people per study. We used a within subjects design due to the time constraint.

System
Building on our two hypotheses we tested three systems. Based upon health literature we used set an explicit goal to walk 150 minutes per week in chunks of 10 minutes.

Ubifit
This system functions the exact same as the Ubifit Garden application for the Nokia phones with one key difference: flowers appear only for walking events. We did not want to introduce confounding factors for other studies in how to feed back different types of events. A flower appears on each walking activity that is longer than 10 minutes.

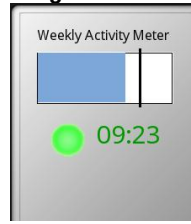Figure 1: Ubifit widget screen capture

Widget

Figure 2: Widget (meter) screen capture

This system feeds back walking distance through a meter that fills up as an individual walks. The stoplight indicates the level of the current activity. A timer appears when the individual is engaged in a brisk walk to encourage the user to complete the current activity for ten minutes so they do not feel a sense of loss-aversion [2].

WidgetR
This system has the same homescreen widget as the Widget system but has an integrated reminder system in the application to input when to exercise. When this reminder is due, an individual has the options to say the activity has been completed, decline the activity or to postpone the activity.

DATA ANALYSIS
In the following analysis, we use the number of walks in a day to quantitatively analyze each group of participants. Due to the small sample size in length of study and large variance in user habits, extrapolating walking data to find significance was unfruitful. The more insightful qualitative feedback from this pilot study is presented in the discussion section.

For each element: $n = 4$, degree of freedom $= 2*n - 2 = 6$

Test	T-value	Two-tailed P-value
H1: Ubifit versus Widget	0.314	0.764
H2: Ubifit versus Widget+R	0.816	0.445

Table 1: Significance values for each population over the course of a week.

Unfortunately, these comparisons of experimental data were not statistically significant. The large variance arose from different walking habits for 12 individuals and a range of phones & phone operating systems. One problem we attempted to fix in the study affected three participants, a feature in the Android operating system disabled activity monitoring in meaningful contexts. This problem is discussed later.

By extrapolating this small range of walking behaviors and data points for a given week, the data analysis reaches a meaningful level of statistical significance when the application records data for over one hundred fifty person-weeks. Furthermore, with a longer term study, statistical trends can be found within a population's ability to maintain and increase physical activity.

This initial pilot study uncovered surprising qualitative feedback that will better refine a larger deployment on the Android Marketplace.

DISCUSSION
Throughout the course the data analysis, three qualitative interviews, and post-study questionnaires, we realized four

high level insights to inform the design of our larger Android Application deployment.

Reflecting back, one participant said:

I felt that the widget didn't report accurately on how much I walk. [...] the way I walk that is very steady and doesn't trigger the accelerometer reading. This is why after day 2 or 3, I start ignoring the meter altogether. If the reporting is accurate, I can imagine myself having more incentive to walk more by seeing the meter.

1. System robustness affects "trust" of application.

This insight can be decomposed into two issues that ran its course throughout our study, both of which affected user perception of the reliability and the accuracy of the system.

The first issue was the CPU usage: when the phone was on using our monitoring service, four users with personal phones experienced major hang-ups in their older Android phones. Due to the multi-threaded nature of Android, these users experienced major lag time and slow down when they used the application. This created a sense of mystery and belief that the application was not adequately recording.

The second issue was an Android operating system "feature" that changed the functionality of a crucial communication channel. In our testing, this affected three of our participant phones and is fixed for all phones with Android 1.5, 1.6, and 2.2. The "feature" prevented the Monitoring Service from running as a background process when the phone screen was off. The users who had this bug had to turn their phone screen on as they walked or exercised to have the phone monitor the activity.

2. Furthermore, participants belief they are more active than the system gives the participant credit. Leading to a "mis-trust" of the system.

Through the post-study questionnaire, almost every individual answered they believed they either had or almost had adequate amounts of exercise in a given week. In the phones which recording was occurring without problems, we received the same input that the participants did not believe they were receiving credit for physical activity. Due to this gap in perception, a participant could check the widget for feedback on current activity level.

In describing the widget, one participant said:

I thought the widget was very helpful. Having it constantly onscreen made me more aware of its presence, and I was more conscious of making sure the phone was with me and application was monitoring whenever [I] was talking. However, I noticed some issues (such as not being able to monitor when the screen was off), so I became more conscious of keeping the screen on whenever I was walking to/from class (which is all the walking I usually did that week).

3. Participants looked at widget for feedback only occasionally.

Essentially, we imagined visualizing both the time and the level of activity would encourage users to complete their most recent ten minute walking activity. However, the ability to get direct feedback on what the current activity level may have been confounded due to the form factor. A feedback we found in interviews was needing to peek at the screen for activity level is not very effective in the course of a walk or a run. Although we had considered the idea of multimodal output (through a sound or image), there is the confounding factor of using more than simply visualization in expressing this sort of ambient information we did not wish to explore.

Finally, the reminder system needed to have a more robust interface to bootstrap an individual. By examining use logs and user interviews, we found diverging uses of it.

4. Reminder system interaction design did not provide adequate guidance for bootstrapping the self-efficacy.

Because of the difficult to navigate interface, this feature went under-used by two of the four participants. Interestingly enough, one participant loved the reminder system while the other did not need it because she constantly walked.

FUTURE WORK
Through the past week, we have been actively developing and testing components for our larger user deployment. The goal has been to improve the aforementioned issues and improve the underlying software engineering that will allow our system to scale upwards for analysis and for data collection.

CONCLUSION
We have rapidly prototyped and user tested for a prototype user deployment for an Android sensing and recording application. By deploying and analyzing user data from a pilot study, we have learned about the way people use our application and a set of design insights moving forward. Applying techniques from health sciences and persuasion technologies to mobile sensing applications, we obtained a set of data analysis that could generalize as we collect more data from more users.

REFERENCES
1. "As Nation Gains, 'Overweight' is relative." CNN. http://www.cnn.com/2009/HEALTH/07/10/attitudes.overweight/index.html

2. Scholl, B. J. (2001). Objects and attention: The state of the art. Cognition, 80(1-2), 1-46.

3. Hutchins, Edwin L.. James D. Hollan, and Donald Norman.Direct manipulation interfaces. (1985)

4. Froehlich J, Findlater L, Landay J. The design of eco-feedback technology. Proceedings of the 28th interna-

tional conference on Human factors in computing systems - CHI '10. 2010:1999.

5. Consolvo, S., et al. Activity sensing in the wild: a field trial of ubifit garden. Proc. CHI'08, 2008, 1797-1806.

6. Costanzo, M., Archer, D., Aronson, E., Pettigrew, T. Energy conservation behavior: The difficult path from information to action. American Psychologist 41, 1986, 521-528.

7. Sunny Consolvo , Predrag Klasnja , David W. McDonald , James A. Landay, Goal-setting considerations for persuasive technologies that encourage physical activity, Proceedings of the 4th International Conference on Persuasive Technology, April 26-29, 2009, Claremont, California

8. Sunny Consolvo , Predrag Klasnja , David W. McDonald , Daniel Avrahami , Jon Froehlich , Louis LeGrand , Ryan Libby , Keith Mosher , James A. Landay, Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays, Proceedings of the 10th international conference on Ubiquitous computing, September 21-24, 2008, Seoul, Korea

The columns on the last page should be of approximately equal length.