InterventionViz: Visual Analysis of Behavior-Change Intervention Dynamics

Category: Research

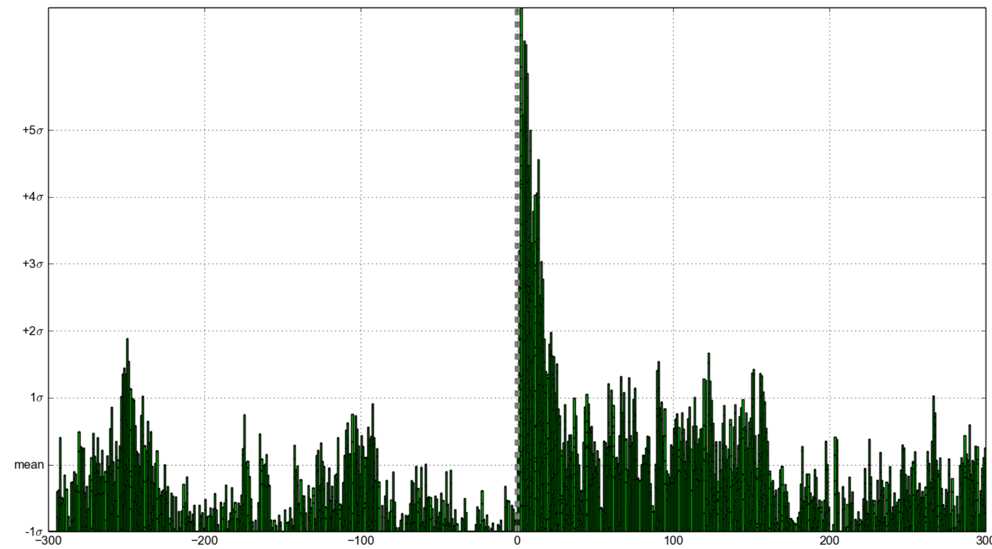

Fig. 1. Sum of step counts following an intervention event shows event response dynamics. TODO: color by events, TODO: also include side-by-side comparison and reference it from text?

Abstract— With the advent of research-grade wearable sensor suites and increasingly ubiquitous human-computer interface comes the opportunity for a new generation of behavioral theory and behavior change methodologies. Just-in-time interventions (JITIs) designed to optimize a subjects behavior based on their momentary context are under exploration as a potentially powerful ally for practitioners and commercial behavior health applications. Analysis of the effect an intervention has on a target behavior, however, is a complex task not well handled by current methods. In this work we explore the application of aggregated time-series visualization techniques to aid analysis of intervention event effects with emphasis on the dynamics of participant response to intervention. To highlight the strengths and caveats of the techniques employed, data from two trial studies of physical activity interventions (n=11, n=10) and one empirical control dataset (n=1) are utilized. The insights presented in this work offer a foundation for future visualization research addressing this problem and as a guide for behavioral scientists in need of more novel methods of analysis.

Index Terms—Radiosity, global illumination, constant time

1 Introduction

1.1 Motivation

1.1.1 Why Health Interventions?

Despite significant advances in the theory and practice of behavioral science, humans continue make poor behavioral choices on a daily basis, and the reasons for those choices remain, to some extent, uncharted. The consequences of these daily choices are often insignificant in the moment but over time build up to larger individual and societal problems. In fact, many of today's greatest health challenges can be mapped back to individual behavioral choices. A habitual lack of physical activity, poor diet, and smoking is likely to lead to a variety of health problems (e.g., obesity, diabetes, heart disease, cancer, chronic pain, depression), lower quality of life and shortened lifespans [6, 3, 20, 16]. Smoking and lack of exercise are behaviors that clearly affect health and longevity (by increasing the chance of getting cancer, or decreasing the chance of obesity and diabetes), but the general populus seems to lack power over their own habits. Similar challenges exist outside the realm of personal health. Academic success is a function of attending class and completing assigned tasks, among other daily behaviors[2]. In personal finance, poor day-to-day purchasing decisions can add up to large financial debts[15].

Behavior change is the art and science of facilitating adoption of new behaviors or abandoning of old behaviors. Behavior change interventions are particularly important to modern health care systems in developed countries, where preventative behavioral treatment recieves little attention from heath care practitioners.

1.1.2 Why Do Intervention Dynamics Matter?

The confluence of pervasive sensing, machine learning, network access, and computation is facilitating new approaches to improving behavioral choices. These systems can detect behaviors and psychological states such as stress[1, 13], physical activity[12, 5], social interaction[19], and smoking[17], automatically and, in some cases, in real-time. These data provide new opportunities for the human-computer interaction (HCI), behavioral science, and other related communities to develop user interfaces for mobile behavioral interventions that help users make better in the moment behavioral choices related to health[11, 14], productivity[9, 18, 10], personal finance[7], and environmental stewardship.[4] New methods for evaluating these behavioral interventions remain underexplored and conventional methods of analysis do not offer the level of detail needed to explore the implicit dynamics of just-in-time, interactive, or adaptive interventions.

Current theories of behavior appear inadequate to inform state-of-the-art intervention development /citeRiley2011, but current methods of intervention analysis are limited in their ability to inform the development of state-of-the-art behavioral theories. In addition to metrics of success, behavioral theorists need tools to help understand the dynamics of behavioral responses to a stimulus. We present methods for exploring these dynamical responses through data visualization.

1.1.3 Contributions

TODO: summarize prev. work (1 paragraph)

TODO: summarize work done (1 paragraph - maybe a long one)

TODO: summarize contributions (1 paragraph - maybe a long one)

TODO: Bulleted list of 2-3 contributions

- to understand how we can use this type of data to accomplish these domain-specific tasks? - to propose new hypotheses and techniques to support these tasks? - to propose lessons learned that may apply for similar viz problems.

TODO: highlight why should viz community care!

1.2 Previous Work

1.2.1 Current Intervention Methods

TODO: EXAMPLES OF JITAI (Just In Time Adaptive Interventions) ANALYSIS and HOW METHODS FALL SHORT

Recent advances in sensing and ubiquitous computing are enabling examination of behavior at much smaller time scales (on the order of seconds) and in a wide range of daily life contexts.

New wearable sensing technologies are changing the way we...

1.2.2 Related Event-based Time Series Visualization

TODO: INSPIRATION TAKEN FROM EXISTING VIZ WORK

2 Design Considerations

TODO: How is the problem here similar to what has come before? How is it different?

A set of design considerations has been developed based on many iterations of user-driven discussions to identify 1) common goals of an intervention efficacy analysis, 2) the key analysis tasks which need to be enabled or improved, 3) typical characteristics of intervention datasets, and 4) weaknesses in current methodologies.

2.1 User Goals

TODO: fill these out

* Characterize how people respond to interventions, especially over time

* Assess intervention effectiveness (efficacy, tolerability, ease of use)

* Characterize how differences among individuals, sub-groups, and contexts affect effectiveness and dynamics of response

* Identify new interventions for future testing (hypothesis generation)

2.2 Tasks

(should follow from goals)

TODO: What types of tasks will they need to do?

see section 4 below

2.3 Characteristics of Intervention Datasets

TODO: What kind of data do they have (or more important starting to get)?

* Multiple time-scales * crossover designs * high-frequency numerical measures: PA, ECG, GPS, ?? * numerical measures with low frequency: EMA constructs, blood-draws * contextual, nominal data: activity classification, location classification, social contexts

TODO: Use descriptions of our datasets to highlight common characteristics?

2.4 Weaknesses in Current Methodology

TODO: What do they need to know about intervention dynamics?

3 Example Application: Physical Activity

As an example application to demonstrate the strengths of the proposed visual analytics we analyze two empirical datasets with a minute-level metric of physical activity level and intervention events delivered throughout a period of ¿8 days. In both studies interventions were delivered with the intent of increasing subjects physical activity, and responses to interventions varied between participants and delivery contexts. In addition to this data, a control dataset with known intervention responses is included for comparison.

TODO: Why these datasets?

TODO: Similarities and differences among datasets.

TODO: do any of the differences suggest different approaches? Do the differences matter for this paper?

3.1 mAvatar Study

An alternating treatment design is used to examine subject behavior over a period of 8+ days in order to test the effect size of an avatar-based live wallpaper deployed on Android phones. Subjects (n=11) aged 11-14 were exposed to a simple, animated cartoon avatar on their mobile device showing alternating levels of physical activity. Each day the avatar would either be active (playing basketball, running, bicycling) or sedentary (watching TV, on a computer, or playing video games). Fitbit One electronic pedometers were used to estimate subject levels of physical activity via step count.

3.2 KnowMe Study

In this study ten teenagers were asked to carry a smartphone and wear an accelerometer and a heart rate monitor (PAdevice??) for 3 days. Physical activity was measured continuously and was monitored in real time. When a subject had been continuously sedentary for two hours, a personalized SMS text message was sent to their phone. Each text message is manually crafted to prompt the subject to be more physically active.

MORE DETAILS NEEDED

3.3 Control Dataset

The control dataset is the result of manual recording of one subject undergoing an imaginary, very potent intervention. Whenever the subject was at his desk a random timer was set for an interval ranging from 10 to 120 minutes. When the timer went off, the time was logged and the subject intentionally increased his level of physical activity for a period of no less than 5 minutes. Step counts were recorded throughout the duration of the two-week study period using a fitbit one electronic pedometer.

4 Tasks

Existing macro-scale methods can determine if an intervention has a significant influence over our target behavior, but it does not give much insight into how the event has an effect over time. The dynamic response to the intervention has only recently become available for study thanks to increasingly ubiquitous wearable sensor technology, and so conventional methods have dealt with low-frequency outcome measures with clever study design. Now that we can measure outcomes at much higher frequencies, methods which leverage this additional information should be adopted.

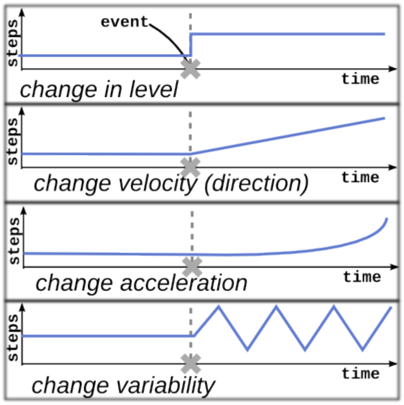

Fig. 2. Theoretical responses to an intervention (adapted from [8]).

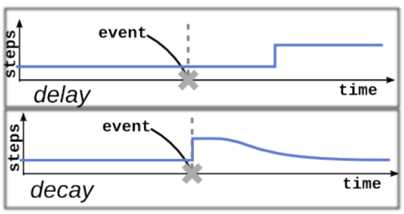

Fig. 3. Level-change responses with delay and decay (adapted from [8]).

The dynamic response of the targeted behavior leading up to and following the event tells us much more about how this effect begins and fades over time. A deeper look into the shape of the signal following our event may even reveal a significant effect overlooked by our previous analysis, and much more quantitative behavioral models become possible.

4.1 Highlighting Event Dynamics

The dynamics surrounding a particular event are most commonly shown using a time series. The instance or span of the event is marked on the time-axis and the value of the behavioral measure (physical activity in this case) is encoded in the height at each point in time. We can describe several idealized intervention types based on behavioral theory using this common visualization paradigm.

Figure 2 shows the case where an event instantaneously causes permanent change in the target behavior, but in the many cases the intervention will have a temporary effect on the target behavior and will have some delay before setting in.

These intervention response dynamics as shown in figure 3 are critically important for just-in-time adaptive intervention developers, but are largely unaddressed in current theory.

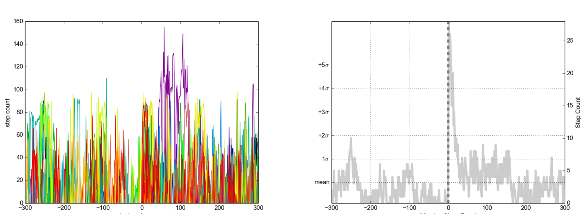

Fig. 4. Comparison of all event responses (left) to average response (right) surrounding the control dataset intervention.

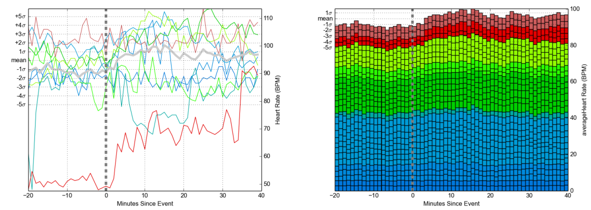

Fig. 5. Comparison of knowMe average lines and stacked bars.

4.2 Event-time Alignment

Figures 2 and 3 give us sense of what an intervention should look like, but in reality individual variations in context completely mask the often small effect of an intervention (see Figure 4 (left)). This is a familiar problem with a familiar solution. By time-shifting our data view so that each intervention event falls at t=0 in a time-series, we can average the series together in order to identify effects which are common between all events. Aggregated measurements surrounding each event begin to give a better picture of the latency and delay of the response.

This approach can be taken for all events in one subject's time series to characterize that subject, or can be applied across subjects to characterize a more generalized response to the intervention.

Figure 4 makes clear the value of this method for drawing general conclusions from a set of seemingly random data. When looking at all events individually (figure 4 left), it is difficult to spot any pattern in the series. One notable event seems to stand above the rest from 20-60min following the intervention, and a cursory glance at this visual might lead one to believe that this intervention was only effective in that one instance. When averaging across all event responses, however, a clear, significant response is evident (figure 4 right).

As expected, figure 4 shows the control intervention to be quite effective at increasing the step count. The additional y-axis showing the mean and standard deviation of the series is included to give an increased sense of the significance of this effect relative to data which may be out of frame. In addition to the nearly immediate response, a longer-lasting effect reaching out to approximately 180m after the event seems to be boosting step count, though the all-events view in figure 4 as well as the stacked-events display in Figure 1 reveal that there are two events which may be the the cause. Though the average line-graph makes spotting effects easy, we must be wary of outlying events or participants which can skew the average. This danger can be sometimes mitigated through use of median in place of mean, but since step-count does not obey a normal distribution (0 values are almost always modal), that approach does not work well in this case.

4.3 Stacking

To address the shortcomings of the aforementioned average-line shown in figure 4, individual events can be shown on the graph and stacked. This yields the same shape, and the y-axis can be easily normalized to match our average series by dividing by the number of events. While still averaging out random contextual influences, this visual also provides indication that the average result is not due to one outlier event, enables easy spotting of missing data or faulty sensors, and gives some indication of the number of events considered. For an n-of-one dataset such as the control dataset, events can be graphed with a unique color. In figure 1, event colors are chosen based on the order in which they were observed. Color mapping of events can also be used to visually group events based on time of day, location of the event, or participant.

Figure 5 shows the difference between a plot of various average response lines and the stacked area plot. The thin lines in figure 5 (left) represent the response of each participant to the event averaged across all events for that participant. The thick line shows the average across all participants average series.

The stacked bars in figure 5 (right) are colored by participant ID, and each represent one unique event - stacked in order of event incidence. This allows researchers to search for both participant outliers within the set as well as event outliers within each participant.

The linegraph allows for identification of unqiue individuals (such as red, light blue TODO: mark these and reference by marker not color), but the stackplot better highlights the overall effect and also shows the number of events considered. Another key difference between these two approaches is the proporitonate weighting of subjects into the global effect display. The line-graph approach considers each subject equally regarless of the number of events recorded of that subject, whereas the stackplot considers the events equally and thus each subject's contribution is weighted by the number of events recorded.

TODO: eqtns

Equation **??** shows the difference in aggregation methods for line vs stacked views.

Figure 5 shows an increase in heart rate (TODO: use a different measure?) following the delivery of a physical-activity-suggesting sms message. Though the behavioral measure differs greatly from that used in the control dataset and 4, a comparison of the y-values in terms of standard deviation also reveals that this effect is less extreme than what we observe in the control intervention.

The deviation from the mean as measured relative to the standard deviation gives a sense of how unlikely the signal is to be a random artifact, but methods for evaluating the statistical likelihood of observing a particular shape are not covered here.

4.4 Controlling for Intervention Delivery Context

In some cases introducing a control event against which to compare the experimental event can help isolate the intervention from the context in which it is delivered. For instance, an intervention delivered on a mobile device is always delivered within the context of phone interaction. That is, the user is always using the phone when the intervention is delivered. It is possible that "using the phone" has it's own unique effect on the behavioral measure. Thus, using "phone use" events as a baseline against which to compare "phone use and intervention delivery" strengthens the chance that the observed effect is a result of the intervention itself and not the result of frequently concurrent contextual forces.

Using the control dataset as an example, random points in time when the subject was at his desk can serve as the control event against which to compare the experimental "intervention" event.

To represent a difference between two intervention effect time series visually we return to the use of line averages due to the confounding nature of negatively-valued bars in stacked graphs.

TODO: fig line graph of controlData avg(intervention)-avg(control)

4.5 Comparing Event Types

Aforementioned methods used to provide a contextual baseline of comparison for events can also be applied to allow for a comparison between two event types. By treating one event as the baseline, differences between the events can be visualized. Using this paradigm, nearly equivalent event responses will have a near-zero difference. Positively-valued areas of the resulting chart indicate times when the "experimental event" had a greater positive effect on the target measure, or, conversely, that the "control event" had a greater negative effect on the target measure.

The mAvatar dataset contains two types of intervention which may be interesting to compare: 1) active-avatar viewing, 2) sedentary-avatar viewing. In this case, the two event types are theoretically opposite in effect, meaning that the sedentary-avatar effect should resemble a mirrored version of the active-avatar effect. Thus, the difference should accentuate the intervention's effect signature and better isolate the behavioral response from noisy data.

Even with two oppositely-polarized events figure 6 fails to show the dramatic effect a researcher might hope for on average. The lack of observed effect indicates that the intervention had no effect on average. In this case, researchers attribute the apparent lack of effect to an ambiguity in study design which led to two opposing conditions: 1) subjects respond positively to physically-active avatars via the Proteus

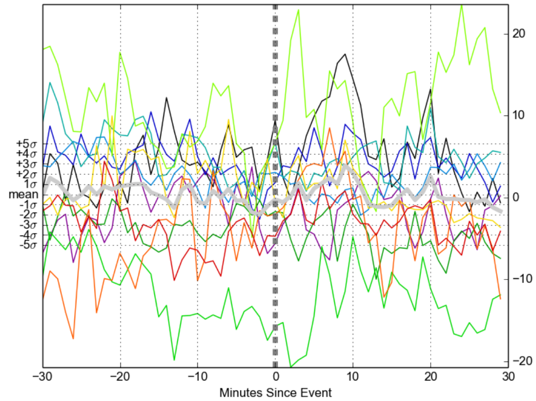

Fig. 6. Active-event series average minus sedentary-event series average. (average shown as bold)

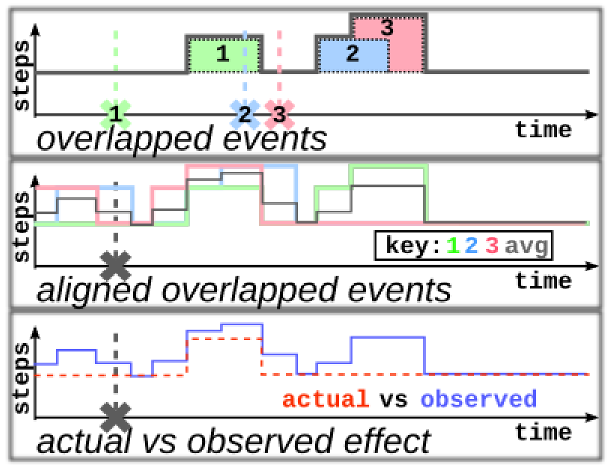

Fig. 7. Overlapping analysis windows confounds observed effect after time-alignment.

Effect [?] 2) subjects respond negatively to physically-active avatars via falsely perceived biofeedback.

Figure 6 suggests this subgrouping within the data in the individual participant series.

TODO: talk about the subgroups (+ response vs -response), label the series.

5 Domain Expert Feedback

TODO: what do domain experts think about these methods (Donna? Eric?)

6 Discussion and Future Work

6.1 Dealing With Overlapped Data Frames

When looking at data surrounding an event, we must be cognizant of how instances of the same event at another time may effect our data. For instance, if our analysis targets the 30min following an event, and the event frequently occurs at 10m intervals, the overlapping signals will create unwanted artifacts.

Figure 7 illustrates this point by showing how events falling within each other's windows of analysis confound the data and ultimately yields a signal quite unlike the actual effect response. In real data, this is further complicated by the non-linear way in which effects are expected to combine. Study designs utilizing methods outlined in this paper should design studies to minimize overlapping analysis windows.

Event overlap becomes somewhat inevitable, however, for large event window sizes.

TODO: figure: comparison of events selected with/without overlap from mAvatar dataset to demonstrate difference (especially at high time intervals like no-overlap for 3hrs around event)

Figure **??** shows a comparison between data retrieved using the two methods of analysis: no overlap vs overlap ignored. Allowing no overlap between events helps ensure that multiple interventions effects do not skew the data, but ignoring these data can drastically reduce data sample if large times following the event are used because very few events are so isolated.

6.2 Modeling Observed Effects

Ever since [], behavioral scientists have sought theories to explain, influence and change human behavior. Theories are built on constructs that explain or act as prerequisites for performance (or lack of performance) of a behavior. For example, constructs such as location, social networks, and time of day (external context) influence behavior as much or more so than a person's internal traits and states, such as personality, emotional state, and mental health. With the right theory in hand, behavioral scientists can develop interventions that target these constructs to effectively drive behavior change (ref to Figure of simple cause and effect model).

However, such theory-driven approaches to behavioral interventions have shown limited success. There are likely many reasons for this, but one clear reason is that the underlying behavioral theories driving such interventions is wrong.

TODO? Some example models of the aforementioned effects for control and/or knowMe data.

transfer function modeling? fitting of a fluid-flow analogy using a priori constructs? other?

6.3 Characterizing Psychological Influence of Events via Response Signature

TODO? different psychological mechanisms of delivery used by an intervention could be identified by the dynamics observed

6.4 Statistical Analysis of Features

TODO: I think that to do this right we need to perform intervention analysis using transfer function modeling within arima framework

7 Conclusion

There are many reasons why behavior change is hard: people build up habits over time, behaviors may be tied to addictive substances or activities, behavior is tied to normal daily human experience and context (i.e. smoking more often when with a certain group of individuals or in a particular location). The full potential of behavioral interventions - particularly JITAIs - will continue unrealized until mechanistic methods of behavior modeling are adopted by the behavior science community.

The presented visualization methods allow for more detailed analysis of how a subjects behavior responds to a stimuli over time. These methods, when combined with a computational modeling approach to understanding human behavior may enable behavioral scientists to formulate more accurate and more application-ready theories, leading to more effective behavioral interventions.

References

[1] K.-h. Chang, D. Fisher, J. Canny, and B. Hartmann. How's my mood and stress?: an efficient speech analysis library for unobtrusive monitoring on mobile phones. In *Proceedings of the 6th International Conference on Body Area Networks*, pages 71–77. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2011.

[2] H. Cooper, J. C. Robinson, and E. A. Patall. Does homework improve academic achievement? a synthesis of research, 1987–2003. *Review of educational research*, 76(1):1–62, 2006.

[3] A. L. Dunn, M. H. Trivedi, and H. A. O'Neal. Physical activity dose–response effects on outcomes of depression and anxiety. *Medicine & Science in Sports & Exercise*, 2001.

[4] J. J. Elliott. Development of an energy-information feedback system for a smartphone application. 2012.

[5] B. A. Emken, M. Li, G. Thatte, S. Lee, M. Annavaram, U. Mitra, S. Narayanan, and D. Spruijt-Metz. Recognition of physical activities in overweight hispanic youth using knowme networks. *Journal of physical activity & health*, 9(3):432, 2012.

[6] O. H. Franco, C. de Laet, A. Peeters, J. Jonker, J. Mackenbach, and W. Nusselder. Effects of physical activity on life expectancy with cardiovascular disease. *Archives of internal medicine*, 165(20):2355–2360, 2005.

[7] D. Gallego Vico, G. Huecas, and J. Salvachúa Rodríguez. Generating context-aware recommendations using banking data in a mobile recommender system. In *ICDS 2012, The Sixth International Conference on Digital Society*, pages 73–78, 2012.

[8] G. V. Glass, V. L. Willson, and J. M. Gottman. *Design and analysis of time-series experiments*, volume 197. Colorado Associated University Press Boulder, 1975.

[9] J. Ho and S. S. Intille. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 909–918. ACM, 2005.

[10] S. Jewell. Productivity via mobile phones: Using smartphones in smart ways. *Journal of Electronic Resources in Medical Libraries*, 8(1):81–86, 2011.

[11] P. Klasnja and W. Pratt. Healthcare in the pocket: Mapping the space of mobile-phone health interventions. *Journal of biomedical informatics*, 45(1):184–198, 2012.

[12] M. Li, V. Rozgic, G. Thatte, S. Lee, B. Emken, M. Annavaram, U. Mitra, D. Spruijt-Metz, and S. Narayanan. Multimodal physical activity recognition by fusing temporal and cepstral information. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 18(4):369–380, 2010.

[13] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 351–360. ACM, 2012.

[14] I. Nahum-Shani, M. Qian, D. Almirall, W. E. Pelham, B. Gnagy, G. A. Fabiano, J. G. Waxmonsky, J. Yu, and S. A. Murphy. Q-learning: A data analysis method for constructing adaptive interventions. *Psychological methods*, 17(4):478, 2012.

[15] J. M. Norvilitis, P. B. Szablicki, and S. D. Wilson. Factors influencing levels of credit-card debt in college students1. *Journal of Applied Social Psychology*, 33(5):935–947, 2003.

[16] R. Ross, D. Dagnone, P. J. Jones, H. Smith, A. Paddags, R. Hudson, and I. Janssen. Reduction in obesity and related comorbid conditions after diet-induced weight loss or exercise-induced weight loss in mena randomized, controlled trial. *Annals of internal medicine*, 133(2):92–103, 2000.

[17] E. Sazonov, K. Metcalfe, P. Lopez-Meyer, and S. Tiffany. Rf hand gesture sensor for monitoring of cigarette smoking. In *Sensing Technology (ICST), 2011 Fifth International Conference on*, pages 426–430. IEEE, 2011.

[18] T. Sohn, K. A. Li, G. Lee, I. Smith, J. Scott, and W. G. Griswold. *Place-its: A study of location-based reminders on mobile phones*. Springer, 2005.

[19] D. Wyatt, T. Choudhury, J. Bilmes, and J. A. Kitts. Inferring colocation and conversation networks from privacy-sensitive audio with implications for computational social science. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(1):7, 2011.

[20] D. G. Yanbaeva, M. A. Dentener, E. C. Creutzberg, G. Wesseling, and E. F. Wouters. Systemic effects of smoking. *Chest Journal*, 131(5):1557–1566, 2007.