ABSTRACT

Wearable computing devices, such as the smartphone, personal digital assistant, and global positioning system have become ubiquitous today. These devices are designed to enhance and enrich daily activities by providing automatic and autonomous support. However, these devices can be intrusive, particularly when the user is immersed in another more important activity. Furthermore, continuous disruptions can often leave the consumer with feelings of frustration or anxiety. One approach to reducing disruptions of wearable devices is to deliver the notifications to the user during times of transition of activity. During activity transition, the user’s perceived burden is lower because they are actively changing their primary task to prepare for another. This research effectively detects physical transition of the user and delivers device-generated notifications during these transitions. An iPhone is augmented with an application that detects physical activity transitions using the data acquired from a built-in LIS302DL tri-axial accelerometer. This work compares user receptivity to messages delivered during activity transitions and of those delivered at random by allowing the user to annotate how receptive they are to each message at time of delivery. The results suggest that context-aware notification mediation can be useful in reduction of the perceived burden of interruptions and, as a result, augmentation of wearables with this type of mediation would be valuable to the consumer.

KEY WORDS

Wearable devices, Handling Interruptions, productivity measures.

1. Introduction

With the increasing popularity of wearable devices, interruptions from these devices generate new demands on the attention of the consumer. Whether it is e-mail, or telephone calls, upcoming meetings notifications, changes in the stock market or navigation directions, notices on a wearable computer can happen anywhere and at any time. Ubiquitous and mobile devices have become integrated into the consumers’ everyday life, as they do enhance and enrich daily activities by providing automatic and autonomous support. For example, a mobile device could provide a daily reminder to take prescription medicine or an convey an announcement of a newborn baby via a text message from a loved one; however, although proactive messages are at times welcomed, there are times when such interruptions are inappropriate or unwanted. For instance, a user receiving a newborn baby announcement during a critical work related task would most likely view the timing of receiving this message as inopportune. Even though most of the devices come with off or inactive option, many of us forget to turn our devices off while working.

Wearable devices proactively communicate through a bombardment of messages thereby interrupting the consumer from his/her present task which is most often independent of the device [1]. Moreover, if several devices access the user, this could lead to multiple and simultaneous interruptions that are insensitive to the user’s workload or context. Studies have also shown that delivering notifications in this proactive manner runs the serious risk of interrupting the user’s ongoing task creating feelings of frustration and anxiety [22].

Because consumers are not always available to receive messages and because ubiquitous and wearable computing devices are pervasive in nature, users perceive these notifications as disruptive. Disruptive notices can lead to feelings of information overload and “interruption irritability” [23]. For that reason, there is a need to minimize the perceived burden of proactive messages with the careful management of notifications. Management of notifications should consider the user’s availability by analyzing their context as well as the message content.

Because wearable devices proactively deliver messages that are meant to enhance and aid our everyday lives but often burden us with unwanted and detrimental interruption, a simple augmentation to the device is needed to allow it to recognize opportune times for communication. Motivated by the observation that a transition between two different physical activities may strongly correlate with switching tasks, a time where it is also noted to be of decreased mental load by Ho and Intille [23,24] and Kern and Schiele [25,26] this research augments an iPhone 3G device with an application that mediates the delivery of non-time critical messages with the detection of four postural activity transitions—sitting to standing, standing to sitting, standing to walking, and walking to standing. A measure of the user’s receptivity
of messages delivered during activity transitions and of those delivered at random is also explored. This work’s main contribution is an iPhone 3G application built for the iPhone 2.2.1 operating system. Experience of small (6) number of users and testing is discussed as well.

2. Related Work

While many of the recent interruption studies concentrate on the personal computer as the producer of interruption [27,28,23,24], studies have also focused on interruption by wearable devices [23-26,11,11]. Disruptive notices from wearables or Personal Computers can lead to feelings of information overload and “interruption irritability” [23,24]. Delivering notifications in a proactive manner often interrupts the user’s ongoing task. For example, studies have shown that interrupting the user’s task at random moments can cause decreased performance on the task at hand as well as feelings of frustration and anxiety [22]. In a social interaction, interruptions delay and distract natural flow of human-to-human conversation and can render actions of people incomprehensible [22]. In addition, an interruption, which is unrelated to the tasks at hand, takes longer to process and return back to original task [27,28]. Although the effects of interruption are more often thought of as negative influence on the work being done [28], there is some evidence that interruption can increase efficiency, productivity, prevent errors, and even influence behaviour if the interruption aids the current task in some way [5,24]. Two key factors contribute to the user’s receptivity of interruption. The first includes the selection of the most appropriate time for the user to receive interruption. The moment chosen to gain the user’s attention can drastically alter the user’s receptiveness towards the interruption [24]. For instance a critical message might be better suited for immediate delivery, whereas a non-time critical message might be better received if it was delayed to a later moment [23,24].

The medium of interruption or the method by which a message is delivered has also been observed to affect the receptivity of interruption [24,28]. For example, consider an office worker engaged in a critical task. He/she might be more receptive to a SMS message alert rather than if the phone were to ring. Because aurally presented interruptions are thought to be acknowledged more quickly than visual stimuli and as a result more resistant to interruptions than visual ones [28], the visual notification is less likely to disrupt the flow of the current task, perhaps lowering the perceived burden of the interruption [24]. Furthermore, thermal interruptions have larger detrimental effect than light on disruptiveness and performance. Additionally, motion as a notification system is effective compared with static items as traveling motions, such as a visual-stimuli, are more disruptive than anchored motions [28].

A breakpoint represents the moment of transition between two observable, meaningful units of task execution reflecting transitions in perception or action [27]. Conversely, the novel work of Wilson and Miller [2,3] in which the observation of natural breakpoints were exploited is explored.

One approach to detect and differentiate breakpoints in user tasks is to use statistical models that map user interaction with physiological measures of the heart’s electrocardiogram and the brain’s electroencephalogram as shown in [1]. Similarly, McCrickard and Chewer [4] use physical or biomedical sensors to infer workload characteristics; yet, they are augmented with sensors that monitor eye gaze to detect breakpoints. Along the same lines, Bailey and Iqbal [22] use an eye tracking system to collect pupil dilation data found as a reliable measure of workload, to align task models for the determination of the best time to deliver notifications. Sawhney and Schmandt [6] use auditory cues of current activities and conversations in the room to determine breakpoints. Horvitz et al. [7,8] monitors the activity of a user interacting with different client devices augmented with event sensing and abstraction which sense computer events from the operating systems and applications executed on their clients accompanied by a microphone to report visual pose and nearby conversation, a camera to capture nearby conversations, and online appointment information to extract meeting properties. With sensor data, a Bayesian head tracking system, audio signal processing, learned models, and decision graphs Horvitz et al. Also, [7,8,11] is able to delay a message during higher mental loads until a user is more susceptible for delivery. Furthermore, Gievska and Sibert [9] created a framework based on an interruption taxonomy and Bayesian Belief Networks learning algorithm to aid interface designers selecting the most appropriate timing for interruption [25,26]. Ho and Intille [24,25] also rely on activity transition to mediate interruptions; however, Ho and Intille use accelerometers as its sole sensor to classify activity transitions and a C4.5 decision tree learning algorithm to deliver messages during transition. Our work also uses the accelerometer to identify the activity transition. Alternatively, rather than attempting to build a model of the user’s mental workload, slow-growth notification—a novel method of notifying users through the use of gradual changes in the visual representation of a message, relies on the user’s own internal interruption model. Because the point at which the user tends to notice the slow-growth notification corresponds to a time at which they are more ready to be interrupted, Wilson and Miller [2,3] exploit these natural breakpoints to reduce the perceived burden of interruption. This work employs techniques presented by Ho and Intille [23, 24], and Kern and Schiele [25,26] using activity transition as an opportune time to deliver messages to the user. Similar to Ho and Intille [23,24] used two accelerometers in their study, one ankle mounted and the other thigh mounted.

3. Design
Together x, y and z form a 3 Dimensional acceleration vector that indicates the direction of gravity [13]. By determining the orientation of the device with respect to gravity, the tilt of the iPhone may be determined. Karantonis et al. [14] uses a single waist-mounted triaxial accelerometer to recognize the postural orientation of the wearer or the relative tilt of the body in space. Karantonis et al. [14] found that tilt angle, the angle between the positive z-axis and the gravitational vector g can be used to recognize the postures such as lying, sitting, upright and standing. As the device used in our experiment (iPhone™) is different than that used in [14], the tile-angle here is the negative of y-gravitational component.

To determine periods of activity and rest, signal magnitude area (SMA) algorithm is used [14-16]. Here, this calculation is performed by summing each sampled x, y, and z values over the time elapsed on one second. Based on the studies in [14-16], a threshold value of .3 is used in our experiments. This value is also exposed to the user for personal changes they might have in mind. A decision tree learning algorithm is used to classify activity transition using the aforementioned sensor data, the determination of postural orientation, and distinction of periods of activity and rest. The decision tree algorithm depicted in [17] accompanied by gathered accelerometer values determines the moment of physical activity transition calculated by the iPhone application by following the leaves of the tree to reach a decision. In this case, a yes or no answer to the question, am I in a state of transition?

3. Implementation

To develop applications for the iPhone OS 2.2.1, an Intel-based Mac running Mac OS X Leopard version 10.5.4 or later is required [13,18,19]. We use a Mac Book Pro with a 2.16 GHz Intel Core 2 Duo running Mac OS X, Leopard version 10.5.6. To perform experiments and test
hypotheses, an iPhone 3G running the iPhone OS 2.2.1 was also purchased. In addition, knowledge of iPhone™ SDK, Xcode™, Objective-C, iPhone Developers™ portal were [10,12] gained before iPhone™ OS Application was developed.

is a set of styles and page layout settings that determine the appearance of a document. This template matches the printer settings that will be used in the proceeding and the CD-Rom. Use of the template is mandatory. A cohesive iPhone application was created to test our hypothesis that activity transitions are an opportune time to deliver device generated interruptions [20,21]. The section is divided into subsections describing the following objectives necessary to carry out our task:

1. Use the built-in triaxial accelerometer to determine activity transitions
2. Generate, Store, and deliver simulated notifications randomly
3. Allow the user to view messages
4. Allow the user to notate their receptivity to the message
5. Allow application settings to be user configurable

Allow results to be retrieved from the device for analysis as explained in [20].

Figure 3(a): Survey View Controller (topleft). (b) Application Setting (top-right), (c) Working and user selected values.

4. Results: Discussion

The performance of the activity detection algorithm was measured against 240 physical activity transitions. The strategy used to validate the activity transition was user annotated activity transitions with the comparison to the results logged from transferring the data from the device to the web server. In a controlled environment, a set of actions were performed to calculate the predictive accuracy of the physical activity classifier that was using a SMA threshold of 0.3 [16]. An SMA value of 0.3 was chosen by consulting the study of [16] in which a threshold value of 3.0 was optimal producing a sensitivity of 0.99 and a specificity of 0.94 in a controlled experiment. The user was instructed to carry out four activity transitions of sitting to standing (SIST), standing to sitting (STSI), walking to standing (WAST), and standing to walking (STWA) by performing a sequence of routine actions as follows:

1. Sit for 5 seconds, stand for 5 seconds, walk for 5 seconds, stop for 5 seconds, and sit for 5 seconds. (SIST, STWA, WAST, STSI)
2. Stand for 5 seconds, walk for 5 seconds, stop for 5 seconds, sit for 5 seconds, and stand for 5 seconds. (STWA, WAST, STSI, SIST)
3. Walk for 5 seconds, stop for 5 seconds, sit for 5 seconds, stand for 5 seconds, walk for 5 seconds, and stand for 5 seconds. (WAST, STSI, SIST, STWA)
4. Stand for 5 seconds, sit for 5 seconds, stand for 5 seconds, walk for 5 seconds, and stand for 5 seconds. (STWA, WAST, STSI, SIST)
5. Repeat Steps 1-4, four times.
6. Stand 5 seconds, sit 5 seconds, and stand 5 seconds.
7. Repeat Step 6, nine times. (STSI, SIST)
8. Walk 5 seconds, stand 5 seconds, and walk 5 seconds.
9. Repeat Step 7, nine times. (WAST, STWA)
10. Sit 5 seconds, stand 5 seconds, and sit 5 seconds.
11. Repeat Step 10, nine times. (SIST, STSI)
12. Stand 5 seconds, and walk 5 seconds, and Stand 5 seconds.
13. Repeat Step 12, nine times (STWA, WAST)

Each set of transitions were annotated a trial, i.e. AT_Trial_1, and by using the iPhone application settings to set the username reflected as such with a total of 28 trials performed. After each trial, the activity transitions detected by the device are sent to the web server. An evaluation of the test results involved comparing the subject’s actual movements with the movements classified by the device. If the transitions reported by the device did not match the actions the user carried out, the transitions that were not reported were marked as missed and the transitions that were incorrectly classified were also marked as such. If at any point, the iPhone application detected a transition outside of trials that is also marked; however, during the trials the iPhone application did not report phantom transitions. Refer to Appendix A for full details of the 28 trials.

In order to measure the predictive accuracy of the activity classifier of the iPhone application, the metrics of precision and recall were used. To calculate precision and recall, the following sets are desired:

- \( S_+ \): denotes the number of positive examples in the test set that are correctly classified positive (A)
- \( S_- \): denotes the number of positive examples in the test set that were incorrectly classified positive (B)
- \( S_\cdot \): denotes the number of negative examples in the test set that are correctly classified negative (C)


- $S_-$ denotes the number of negative examples that were incorrectly classified positive (D)

Both set precision ($A/A+D$) and set recall ($A/A+B$) can be calculated for experiments we conducted. Although, usually one only considers the positive examples of precision and recall formulated by the calculations above, the negative examples can be obtained in a similar fashion. For four activity transitions SIST, STSI, WAST, and STWA, precision and recall are calculated by obtaining $S_+$, $S_-$, $S_+$, and $S_-$. With a set of SIST transitions, $S_-$ is the number of times a SIST transition was detected by the iPhone application as the user performed a SIST transition, $S_+$ is the total number of SIST activity transitions that were reported but not performed by the user. $S_-$ is the number of activity SIST transitions that were performed by the user but not reported as a SIST activity transition, and $S_+$ is the number of times the iPhone application correctly did not detect a transition SIST activity transition. Because for each trial there was five second duration for all activities and because the classifier attempts to determine activity transition each second, $S_+$ is calculated by the multiply of SIST activity duration with the number of SIST activity transition trials subtracted by the number of activity transitions detected for the duration. $S_-$, $S_+$, $S_-$, and $S_+$ values for STSI, WAST, and STWA are calculated similarly.

Table 1 contains a summary of the results. As the above table suggests, a SMA threshold of 0.3 produces fine results providing precision values of 93.33% and 95% for SIST and STWA, WAST, and STSI respectively. Recall values are likewise admirable results of 100% for SIST and STWA, 93.44% for WAST, and 98.25% for STSI transitions.

Because the trials were only performed once for one individual, precision and recall curves cannot be generated. In order to do this, we believe that an automated way to gather metrics must be created because hand calculation in imperfect and takes too much time. Instead, a Receiver Operator Characteristic Curve (ROC) is created by calculating each NER classifier’s true positive rate (TPR) and false positive rate (FPR) and plotting TPR vs. FPR. TPR is calculated same as recall. And FPR is $(C/C+D)$.

**User Experiment:** To test the burden of interruption by the augmented iPhone, six colleagues were asked to participate in an informal experiment. The experiment consisted of two groups of participants, those in the control group and those actively receiving messages only during activity transition. The duration of the experiment was one hour. All participants were told to go about their activities as normal were unaware of the hypothesis of the study and told that the application was going to tell them when it is a good time to check their messages. The participants were then told to rate their receptivity to each messages given to them by the device. Refer to Appendix B in [20] for additional details.

**Device Placement:** The iPhone was clipped, using a belt clip, to the participants left pocket orienting the device facing up with screen accessible to the user.

<table>
<thead>
<tr>
<th>Activity Transition</th>
<th>S ++</th>
<th>S +−</th>
<th>S −+</th>
<th>S −−</th>
<th>Total True+</th>
<th>Total True−</th>
<th>Precision</th>
<th>Recall</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIST</td>
<td>5</td>
<td>0</td>
<td>60</td>
<td>36</td>
<td>2</td>
<td>24</td>
<td>29</td>
<td>93.33%</td>
<td>0%</td>
<td>98.3%</td>
</tr>
<tr>
<td>STWA</td>
<td>5</td>
<td>0</td>
<td>60</td>
<td>37</td>
<td>2</td>
<td>24</td>
<td>29</td>
<td>95.00%</td>
<td>0%</td>
<td>98.7%</td>
</tr>
<tr>
<td>WAST</td>
<td>7</td>
<td>4</td>
<td>60</td>
<td>34</td>
<td>2</td>
<td>24</td>
<td>29</td>
<td>95.00%</td>
<td>7%</td>
<td>98.7%</td>
</tr>
<tr>
<td>STSI</td>
<td>7</td>
<td>1</td>
<td>60</td>
<td>31</td>
<td>2</td>
<td>24</td>
<td>29</td>
<td>95.00%</td>
<td>7%</td>
<td>98.7%</td>
</tr>
</tbody>
</table>

Table 1: Recall and Precision summary results.

The following table and figure presents the results:

<table>
<thead>
<tr>
<th>Activity Transition</th>
<th>TPR</th>
<th>FPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIST</td>
<td>100.00%</td>
<td>1.6</td>
</tr>
<tr>
<td>STWA</td>
<td>100.00%</td>
<td>5.0</td>
</tr>
<tr>
<td>WAST</td>
<td>93.4%</td>
<td>4.2</td>
</tr>
<tr>
<td>STSI</td>
<td>98.2%</td>
<td>8.0</td>
</tr>
</tbody>
</table>

Table 2: Activity transition TPR and FPR

All participants were given the same instructions. In addition, participants were also given verbal instructions not to turn off the application at any time during the hour and to treat incoming messages without context and as non-time critical. All participants seemed to grasp the instructions with ease [20].

**Control vs AT group:** Three of six participants were sent random messages during the hour and three were sent messages only during physical activity detection. All participants were unaware which group they were placed in so as not to affect the receptivity of messages. All
participants were between the ages of 23 and 45, healthy, and work on a computer at a desk during work hours.

**Participant’s interview:** After the experiment, the participant was asked to fill out an assessment of their experience. The following questions were asked of the participant:

1. Is this your first time interacting with an iPhone?
2. Do you find using the iPhone difficult? Did you find using the iPhone application during testing difficult? Please describe.
3. Did anything out of the ordinary happen during the test?
4. How many times did the iPhone interrupt you today?
5. Did you find yourself becoming more or less receptive as time progressed?
6. Can you put in your own words how you were interpreting each of the receptivity numbers, and what it meant in terms of your interruptibility?
7. How much time would it take to respond to each interruption typically (in minutes)?
8. Can you think of a different way that it could have interrupted you so that you would have been more receptive?
9. In the situation where you answered a 1 (Not at All Receptive), do you think you would have been more receptive to an interruption of a different type (i.e. not Phone or Text)?
10. Was wearing the iPhone on the belt clip cumbersome? How can we improve your comfort level?
11. Did you notice any pattern behind the interruptions?
12. How did people around you respond to the system when an interruption occurred?
13. Do you have any questions or comments on the test?

Questions 1 and 2 were asked to get a feel of how interacting with the iPhone and an application on the iPhone was to users who already knew how to use an iPhone and those who did not. Question 3-7 are used for the interpretation of the survey results. Question 8, 10, 12 and 13 is used for the possible improvement of the iPhone application and its comfort level. Question 11 is used to determine whether or not the participant noticed the point of interruption is during activity transitions or is completely random. Additional questions were asked by the administrator of the test to interpret a receptivity of messages.

**Lickert Scale of Receptivity:** We use a Likert scale as a measure of receptivity of messages because we are asking the participant their feeling or attitude about how irritated or receptive they are to receiving a message at the time of receipt. Given that a Likert scale is inherently an imperfect due to its inherent biases, respondents may avoid using extreme response categories, (central tendency bias); agree with statements as presented (acquiescence bias); or try to portray themselves or their organization in a more favorable light, (social desirability bias) Because of the small sample size, we cannot state any conclusions about the responses with statistical significance. However, we can deduce improvements to the iPhone application and testing methods with the following analysis of user responses.

**Activity Transition Group:** Participant A, B and C were in the activity transition group. All receptivity answers were grouped by experiment group and averaged to get individual and group values as seen in Table 3. Table 3 depicts the participant responses for the experiment. Participant A had an average receptivity score of 4.2 but was less receptive to messages during STWA transitions bringing down her average response. 4.2 is somewhere between Mostly Receptive and Most Receptive and a seemly good indication that participant A is Mostly to Most receptive during activity transition. When asked why a low rating was given at this time, the participant noted irritation was due to how quick the interruption was after a previous one. In fact, a STWA transition would have been more received if it hadn’t just come within minutes of a previous transition.

<table>
<thead>
<tr>
<th>User</th>
<th>Not at All Receptive</th>
<th>Less Receptive</th>
<th>Some what Receptive</th>
<th>Mostly Receptive</th>
<th>Most Receptive</th>
<th>Average Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td>4.1875</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3.2222</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1.4444</td>
</tr>
<tr>
<td>A II</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>5</td>
<td>9</td>
<td>3.6333</td>
</tr>
</tbody>
</table>

Table 3: Receptivity Responded

<table>
<thead>
<tr>
<th>User</th>
<th>Not at All Receptive</th>
<th>Less Receptive</th>
<th>Some what Receptive</th>
<th>Mostly Receptive</th>
<th>Most Receptive</th>
<th>Average Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>5.0000</td>
</tr>
</tbody>
</table>
Participant B noted similar irritation. At first glance of participant B’s receptivity answers with an average of 3.2, it seems that participant B is only Somewhat Receptive during activity transition. However, after the wrap-up interview low receptivity was due to again receiving messages within minutes of each other. This leads us to conclude that perhaps the receptivity of notification mediation would improve if there was an inherent delay between deliveries of messages even if there are detected activity transitions. Yet, receiving a message so close to when the messages are checked is unlikely to happen in real life and the application would not interrupt the user without a stored message to send. Nonetheless, consideration to improve the iPhone application is given.

Participant C seems to invalidate our premise of notification mediation and delivery message only during activity transition will reduce the perception of burden of the user. **Control Group**: (Table 4) In addition to the activity transition group, individuals were also tested by random interruption. Participant D, E and F were in the control group. Participant D had a high receptivity value of 5.0 leading to conclude that random interruption is well received.

Participant E, on the other hand, had seemingly very high receptivity to all random messages but the user interview did not match survey results. When interviewed, participant E indicated a strong irritation to receiving a message during a phone call but marked the irritation as Mostly Receptive. When asked about how the receptivity scale was interpreted, participant E understood the correct difference between Mostly Receptive (4) and Not at All Receptive (1). Was participant E trying to be nice? We believe greater numbers in testing would alleviate any confusion on such a discrepancy. Participant E also leads us to believe that perhaps a better scale in testing the device is desired. Conceivably a yes or no answer to, *Are you available to take a text or phone message at this time?* Additionally, augmenting the device to allow the user to set when the next time to ask if the user is available similar to a Calendar application that allows the user to set when to be next reminded of an appointment could improve overall receptivity of device generated interruptions.

Conversely, Participant E produced results closer to those expected with a receptivity average of 3.8, somewhere between Somewhat Receptive and Most Receptive. Mimicking real life, there were times when the user was available to be interrupted and there were times were she was unavailable to be interrupted. When asked why low receptivity answers were given, participant E noted that interruption occurred when busy with a task and receptivity was lower when messages were delivered close together as also noted by participants A and B.

### 5. Conclusions and Future Research

We feel that this small user experiment was beneficial to the understanding of how users will interact with the device, what improvements can be made to testing the application, and what improvements can be made to the iPhone application. Testing should be performed on a larger sample size, for much longer duration, and with a more interpretive scale of interruptability of the device at the time of delivery. Furthermore, perhaps additional smaller tests rather than a more formal large scale test are preferred to work out additional kinks in the system. Also, a more cut and dry, yes-no, answer would perhaps give greater insight to the right timing of interruption than a Likert scale with inherent bias.

Improvements to the iPhone application could also be made. Buttons need to be bigger with more readable text. In addition, a more audible form of interruption is desired. Most participants did not feel the device vibrating. This could also be due to the placement of the iPhoneOne participant noted that they would rather wear the device horizontally on their belt. This could be achieved with a minor tweak to the tilt detection algorithm. First the algorithm could check the orientation of the device, and then calculate the orientation of the user with respect to gravity. Furthermore, incorporating a delay after activity transition detection and allowing the user to set delay user notification reminders is desirable.

In addition to the improvements discussed in the previous section, this work could benefit from additional sensors and methods that detect when the user is available to receive interruption. If the application took advantage of GPS and Calendar events, specific locations or appointments could be tagged with *do not disturb* or *freely disturb* status. The application would then sense user preference when in a specific location or during an event and act accordingly. This seems like the best solution—providing control to the user, accept that an emergency must override the *do not disturb* set up by the user.

Furthermore, the context of the message could be of great importance to the user’s perceived burden of interruption. For example, if a message is sent by the user’s boss or marked as urgent by a friend, the user would want to receive those messages immediately.

This work would very also benefit from a more formal interruption study with a diverse set of participants with varied age and occupation. Perhaps notifications during physical activity transition are only well received to those

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.6667</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.8000</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>4.3636</td>
</tr>
</tbody>
</table>

**Table 4: Receptivity Responses.**
with more sedentary jobs. The device could avoid delivery messages during activity transition or when detecting a large quantity of transition during a short period of time, continue to hold messages only to delivering messages during infrequent activity transitions.

References
[20] First Author, Second Author, XXXX, pp. 1-124, University ......