Exploring Evaluation Methods for Ambient Information Systems

William R. Hazlewood
School of Informatics
Indiana University
901 E. 10th St.
Bloomington, IN 47408, USA
whazlewo@indiana.edu

Youn-kyung Lim
School of Informatics
Indiana University
901 E. 10th St.
Bloomington, IN 47408, USA
kmakice@indiana.edu

Kay Connelly
School of Informatics
Indiana University
901 E. 10th St.
Bloomington, IN 47408, USA
connelly@indiana.edu

Kevin Makice
School of Informatics
Indiana University
901 E. 10th St.
Bloomington, IN 47408, USA
kmakice@indiana.edu

CHI 2008, Florence, Italy
ACM 1-xxxxxxxxxxxxxxxxx.

Abstract
In this paper, we begin by laying out our motivation for exploring methods of evaluating Ambient Information Systems, with a strong push toward in-situ studies. Next, we describe a simple study which was conducted to give us further insight into this research domain. We conclude by discussing the insights gained from our study, and possible ways to improve our evaluation results in future iterations.

Keywords
Ambient Information Systems, Peripheral Displays, Evaluation Methods, Ambient Orb

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
Over the last decade, the advent of large scale ubiquitous computing research has produced an interest in alternative displays that can provide useful information while blending smoothly into the surrounding environment. These devices are distinguished from more typical informational displays in that they are designed to be minimally attended and perceivable primarily from outside of a person's direct focus of attention, providing pre-attentive processing without being unnecessarily distracting. These technologies are intended to be embedded in existing environments, often making use of unused physical and visual aspects of everyday objects. These technologies provide an information channel that can be easily ignored when there are more important matters that require one's attention [5].

Because of their ability to be easily "tuned out," informational displays such as these could offer a way to keep people highly informed while avoiding some of the possible side-effects of ubiquitous technology, such as information overload [8].

Ambient Information System

Researchers have used a variety of different terms to describe their implementations for this category of display. These include peripheral displays, ubiquitous technology, informative art, everyday computing, glancable displays, notification systems, and slow technology. Pousman and Stasko proposed the term *ambient information system* (AIS) as a means to describe all such technologies [6]. This term describes any technologies that:

- Display information that is important, but not critical
- Can move from the periphery to the focus of attention
- Focus on tangible representations in the environment
- Provide subtle changes to reflect updates in information
- Are aesthetically pleasing

UbiComp research has produced several interesting AIS implementations, but work is currently lacking to describe how people's experience with these technologies can be deeply understood. There have been relatively few long term, in-situ, or longitudinal studies conducted to expose optimal design patterns for user acceptance, or show us whether we can accurately say that these technologies are presenting information in a manner that is beneficial in the long run. Some exceptions include Stasko's 2005 study where highly personalized information displays were observed in participant offices [7]; and work from Hsieh and Mankoff two years earlier used an intersting usability/distraction/awareness framework to evaluate an AIS developed to filter important email [3].

Most recently, Tera Matthew's thesis [4] provides a great deal of useful information regarding the design and evaluation of AIS. However, this area of study is still hindered from a general lack of research.

This academic void is partly due to the difficulty involved in creating user studies for this class of technology [2,7]. Because AIS are designed to be highly subtle, and used indirectly, traditional user interface evaluation methodologies can only tell us about the *quality* of a specific implementation (could it work), but little about how these technologies function in the world (do they work).

This is not to say that typical usability methods are not important for AIS. Usability tells us whether or not certain features are working—for example, whether a person is able to detect subtle changes in movement, shape, or color, of a particular design. Unlike most technologies, AIS do not function as intended until they have properly blended into the fabric of the observer's everyday environment. Integrating a new object into the envrionment is a process which requires a certain amount of time, and cannot be accurately simulated inside a controlled laboratory, which is outside the observer's everyday environment. This is why different frameworks and evaluation methods that focus on studies conducted *outside* the laboratory have to be considered to gain a deeper knowledge of how AIS function within actual contexts of use.

Studying AIS in Context

In an HCI laboratory, any kind of observation effects what is being measured. Researchers are often able to account for the effects artificial environments and tasks by separating out artifacts of the lab setting from the valuable bits of information that are independent of the simulated environment. With AIS technologies, however, we cannot understand their effects until we can be sure that the observer is in no way perceiving them directly.

Figure 1. *The Ambient Orb from Ambient Devices consists of a small frosted glass sphere that can shift between thousands of colors, and is controlled wirelessly by a cellular pager network.*

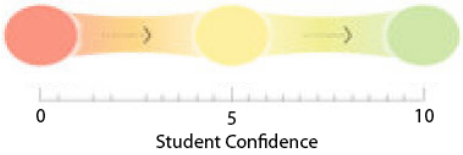

Figure 2*. Color mapping from red to yellow to green, used to display the student confidence level to the instructors.*

Typically, lab studies require that a participant be introduced to the AIS, and then asked to conduct some sort of task which would distract them enough to *force* the AIS out of their primary attention. Even if this study were to run over a long period of time, any probing from the researcher acts to pull the AIS back to the foreground of attention. The ability of such a study to give insight as to how AIS will integrate into the environment is debatable, since having the distractions of a task obscure an information channel is not necessarily the same thing as having that information channel exist within the periphery.

The only way to understand how people will make use of AIS is by conducting long-term in-situ studies, where the participant is allowed to own the technology and even forget about it. Researchers return to probe at a later date. The dificulty is in finding a solid means of collecting data without without distrubing the highly delicate relationship between the participant and his or her peripheral awareness.

Creating an Initial Study Design

Our overall goal is to explore possible evaluation methods to find out what is most useful for AIS considering the concerns discussed above. To explore evaluation methods, we needed to start with a simple exploratory study. Based upon our initial discussions, we developed a set of four criteria we felt were necessary to address our concerns. Our study must:

(i) Make use of a simple existing technology.
We did not want to complicate our study with highly abstract mappings of symbols to meanings, or include anything that would require significant training to be clearly understood. Also, we want to use a system that is somewhat representative of the existing technologies in this domain. That is, our foucs is not on creating a novel AIS, but to look at evaluation methodologies for this class of technology.

(ii) Provide an information source that is of particular interest to the participants.
The information channels used for ambient displays typically include sources that can be constructed from readily available data streams. Examples include: stock information, weather, price alerts, pollen levels, and local traffic congestion. To create a system that provided information of higher relevance to the observer, we opted to create a custom information channel, tailored to our participants' specific interests in a manner similar to Stasko [7].

(iii) Be conducted outside of a laboratory setting.
As discussed earlier, in order see the real effect of an AIS implementation, it is essential that it be observed in an actual context of use.

(iv) Run for an extended time.
To overcome any initial novelty effect that could keep our AIS from becoming fully embedded into the environment, we wanted participants to use the AIS for as long as possible so that the novelty could wear off.

To satisfy (i), we choose, to use the Ambient Orb (Fig. 1) coupled with a simple informational webpage (Fig. 3) that shows a detailed explanation of the orb's state. The Ambient Orb has been used to display a variety of different information sources, and it is a technology that has been eveluated in other HCI research [3].

For (ii), we choose university instructors as our participants. We created an information channel that allowed students to provide simple feedback regarding their courses. That feedback scaled from 1 (low) to 10 (high) and mapped to colors on the orb, ranging from red (low) to yellow (medium) to green (high) (see Fig. 2). We configured the orb to show the mean score reported by the students after each class met.

For (iii), we gave the Ambient Orbs to our participants, who were told that they could feel free to situate them

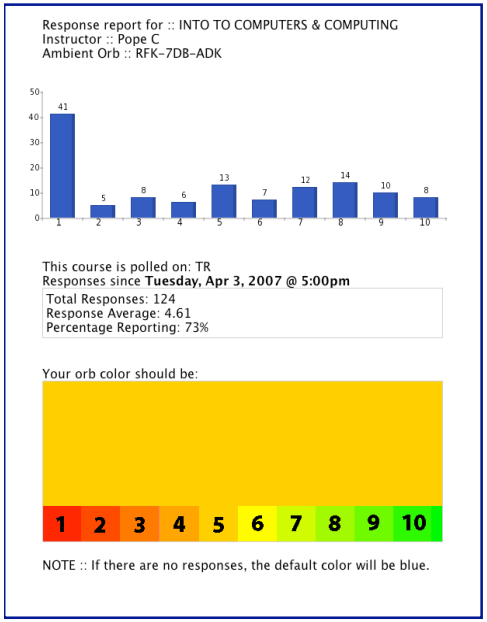

Figure 3. *The detailed webpage showing each instructor the responses received on the classroom poll, and how it affected the orb's color.*

	1st 34 days	2nd 34 days
Instructor 1	1	2
Instructor 2	8	3
Instructor 3	18	21
Instructor 4	12	8
Instructor 5	4	3
Instructor 6	1	3

Table 1. *Total number of instructor detail page hits (1st 34 days and 2nd 34 days)*

wherever they felt it was most convenient. They could even take the device home if they wished. Finally, for (iv) we structured the system so that it could run automatically and log information for as long as the study would allow.

Study Setup

Hsieh and Mankoff conducted a field study where participants were exposed to an ambient system for two weeks [3]. During that time, the participants would receive several pop-up windows on their machines throughout the day with a short survey. An even longer study is in Stasko, J., et al [7], in which observation lasted four weeks in total. The evaluation method for this study consisted of an initial consultation with the participants, deployment of the display, two weeks of use followed by a structured interview and survey, and another two weeks of use followed by the same interview and survey.

When considering these two studies, we were concerned that, while there were periods of use with no distraction, particularly in Stasko's work [7], the amount of time still might not be sufficient to allow the devices to become part of the participant's environment. Based on this, we decided to run a study for at least two months.

We considered probing during the study, as Hsieh and Mankoff did [3], but this clearly trains the user to attend to the ambient device where they might not otherwise. Instead, we devised the system such that, if the users are feel prompted to look deeper into the details, they can refer to an informational web page (not distracting, easy to log). Our hope was that we would see a resulting page access when certain changes occurred in the display. If participants check the detailed information after a change in the display, then we can assume they noticed the display.

Six instructors were gathered from four different departments in the university to participate in this study. Each of the instructors were met individually, given an orb, and briefly instructed on how it functioned. The instructors were told how to know if the orb is working correctly and that they should be able to place it anywhere they desired.

We informed each of the instructor's students that they would be receiving an automated email after each class met, and that each time they would be asked to answer the same query:

"Please rate your confidence regarding your understanding of the materials covered in class today from 1(low) to 10(high)."

To respond, the students only had to click on one of the hyperlinks embedded inside of the email.

Observations from our Initial Study

We were able to run our study for 68 days without directly interacting with the instructors. Upon analyzing the logs from our webserver, we found that the participants made very little use of the informational website, which meant we could not make any strong inferences. Also, the student participation was highly inconistant, which may elude to a weakness in the design of our AIS that could cause errors in evaluation.

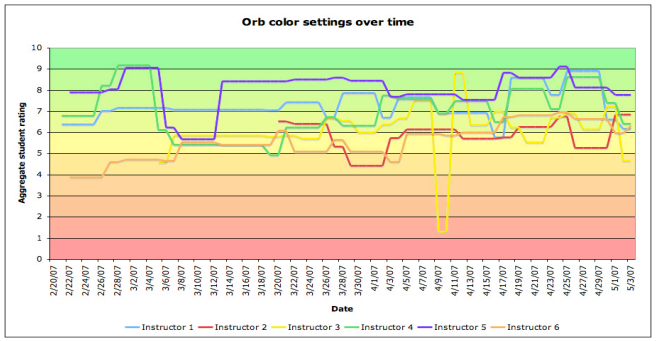

Figure 4*. A graph showing the color value associated with each user's orb (y axis), over the duration of the study (x axis). The color of each bar along the y axis (1-10) shows the colors the orb could range from (Red to Yellow to Green). For most of the participants, the orb changed very little over the entire study.*

As can be seen in Table 1, most of the participants visited the website only a few times. When we asked about this during the structured interview, one participant replied:

"I'd guess I've checked it about 5 or 6 times. I would have checked it more often had I had more change in the color. That would have prompted me to check again. I had such consistency in color that I didn't go look at the web much. When I did it was just because I was curious about the distribution"

We found this sentiment seemed to be consistent among the other five participants. We knew that some of the instructors should have seen at least some change in their orb based upon how the students were responding. After analyzing all of the data that was sent to the orbs, we found that for all participants, the color remained steady for long periods of time (Fig. 3). The lack of variation in the orbs is not necessarily an indication that the system was functioning incorrectly, or that it was poorly designed. During the structured interview, one instructor remarked:

"I wouldn't say the orb was a waste, I used that as a barometer on how frequently to go check the web site. I didn't check the website necessarily a whole lot. I checked the orb a lot, and if I'd seen changes I would have gone to the website."

The orb is reporting the appropriate data based on how the students were responding. Technically, this indicates that the device is fulfilling its role—subtly informing the instructor that there is no need to check the extra information on the informational webpage, since everything seems normal. However, the lack of variance shown on the instructor's orbs, and the resulting lack of activty from the instructors, makes it difficult for us to say whether or not the AIS affected the instructor's envrionment.

The other problem we observed was in the means of collecting data for our custom information channel. While our system was completely automated, we had no control over whether or not the students responded. We assumed that we made the process convenient, but when we checked the percentage of students responding to the automated messages, we found that it was highly erratic. Over the entire study response rates ranged from 12% to 98%.

This lack of consistancy could represent weakness in our system because it lessens the amount of confidence that the instructors have in the data being displayed. That lack of confidence could result in the requirement of extra consideration when the instructor becomes aware of the orb. Instead of being able to simply note "the orb is red" and know what that implies, the instructor also has to wonder, "Why is the orb red?" It could mean that either the entire class is unhappy, or that only the small percentage that responded are unhappy. Design flaws like this will likely have a signifiant effect on the quality of an evaluation because the researcher can no longer rely on his or her assumptons regarding the function of the system.

Discussion and Future Work

The results of our initial study have given insight into the difficulties of developing in-situ studies for AIS technologies. We have found out that even when the system functions sufficiently it is not clear how to assess when it is providing any benefit those perceiving it without using methods that are more distracting. We can only observe and hope for a situation to happen naturally that can be analized. In our set up, the AIS was not a distraction to the instructors, but its not clear whether we can say it provided any benefit.

From the lack of data variance displayed on the orb, it is difficult to know if the orb was non-distracting, or if the instructor simply became unaware of it's presence. To address this distinction it would be useful to know how often, if at all, the instructor made eye contact with the orb. One possible solution would involve video recording the environment where the orb is situated and conducting a video analysis to determine the instructor's awareness of the device. A more promising alternative is embedding eye tracking software into the device so that we have an automated tally of when instructor becomes aware of the device, consciously or unconsciously. This method was used to trigger a modal display in Altosaar's AuraOrb [1], but as of now the particular technology used is not comericially available.

Because of this study, we have begun discussing the feasibility of introducing artificial events in an in-situ study. While technically this method would violate some definitions of in-situ, it would offer an excellent platform to do non-invasive studies on AIS. In this particular study, one possibility would have been to send a false report to the teacher's orb (i.e. forcing it to go red at some point). The obvious problem with this concept is the methodological and ethical issues raised with the introduction of that intervention.

When this study is run again in the spring of 2008, we plan to experiment with different ways to display the responses from the students on the orb to see if we can provide a slightly more useful channel of information without making the display any more invasive. Our hope is that continuing experimentation with our current setup will offer us further insights as to how one can determine the benefit of this particular type of information system.

References

[1] Altosaar M, et al., "AuraOrb: social notification appliance," In Proc. Of CHI 2006, pp. 381-386.

[2] Carter, S. & Mankoff, J. "Challenges for Ubicomp Evaluation." Computer Science Division, University of California, 2004.

[3] Hsieh, G. & Mankoff, J. "A Comparison of Two Peripheral Displays for Monitoring Email: Measuring Usability, Awareness, and Distraction." UC Berkeley, EECS Department. Technical Report No. UCB-CSD-03-1286. 2003.

[4] Matthews, T. "Designing and evaluating glanceable peripheral visualizations." Ph.D. Thesis, EECS Department, Computer Science Division University of California, Berkeley, Technical Report No. EECS-2007-56, 2007.

[5] Olivier et al., "Crossmodal Ambient Displays," People and Computers XX In Proc. of Engage, 2007.

[6] Pousman, Z. & Stasko, J. 2006. "A taxonomy of ambient information systems: four patterns of design." In the Proc. of AVI 2006, 67-74.

[7] Stasko, J., et al. "Evaluating the infocanvas peripheral awareness system: A longitudinal, in situ study." Technical Report GIT-GVU-05-08, Georgia Institute of Technology, 2005.

[8] Wurman, R.S. Information Anxiety. Doubleday, 1989