They want to tell us: Attention-aware Design and Evaluation of Ambient Displays for Learning

Dirk Börner, Marco Kalz and Marcus Specht
Welten Institute, Open University of the Netherlands
Valkenburgerweg 177, 6419AT Heerlen, The Netherlands
{dirk.boerner, marco.kalz, marcus.specht}@ou.nl

ABSTRACT
This paper explores the interaction between users and ambient displays and the evaluation thereof in a learning context. A formative design study examined the user attention towards ambient displays as well as the influence of different display designs. Experimental prototypes were varied on two design dimensions, namely representational fidelity and notification level, and deployed on a university campus. For the evaluation a combined approach using quantitative attention data as well as qualitative assessment methods was used. The results show a high degree of user interest in the displays over time, but do not provide clear evidence that the design of the displays influences the user attention. Nevertheless, the combination of quantitative and qualitative measurement does provide a more holistic view on user attention. The gathered insights can inform future designs and developments of ambient displays also beyond the learning context.

Author Keywords
Ubiquitous computing; public displays; sensors; attention-awareness; contextualisation; design; evaluation.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
More than a decade ago Wisneski et al. envisioned ambient displays as new approach to interfacing people with online digital information [15]. Since then this vision has become concrete. Nowadays, information is conveyed by a plethora of embedded displays and devices, which address all senses and the human peripheral perception capabilities in a subtle or “ambient” manner. On their way to define the design space of such ambient information systems Pousman & Stasko [11] reviewed several definitions and the behavioural characteristics of respective systems. Besides stating that they mostly deal with non-critical information, make use of tangible representations, reflect subtle changes, and are designed aesthetically pleasing, they derived the ability to “move from the periphery to the focus of attention and back again.” The reviewed systems were typically used in divided attention settings with an attention need secondary to a primary task. This notion was also fundamental in Wisneski et al. original work [15]. The authors even described the existence of an attention threshold needed for the transition from background to foreground to take place. Thereby the threshold depends on the user’s current state of attention. Similarly Michels & Müller in their audience funnel framework [9] identified different phases classifying the audience from by-passers over viewers to direct users and allowing in turn to position a display in or in between these phases demanding certain degrees of user attention.

Dealing with or even demanding user attention it is eligible to classify ambient information systems or displays as attention-aware interfaces. In their literature review Wood et al. listed issues to consider when designing such systems [16]. Among others the authors tried to clarify how to measure attention, e.g. by investigating the direction of user gaze. They concluded that there are several limitations on inferring the focus of attention from the direction of gaze and that as an implication the focus of attention should be validated through further evidence, i.e. knowledge transfer. Addressing this recommendation when dealing with ambient display design imposes several questions related to the modelling of user attention towards the display and the evaluation thereof.

A review of existing literature on ambient displays for learning [1] found several design and evaluation guidelines including links regarding the evaluation of user attention and knowledge transfer. Mankoff et al. adapted the heuristic evaluation for ambient displays [5] and called for a “sufficient information design” presenting “useful and relevant information” using a “consistent and intuitive mapping”. McCrickard et al. classified and evaluated notification systems based on the parameters interruption of the primary task, reaction to informational cues, and comprehension of the presented information [8]. Justifying their framework they emphasised the need for a balanced system design and evaluated the trade-off between attention and usability in a series of experiments [7]. An evaluation
framework for the comprehension of ambient displays [4] has also been introduced. The framework distinguishes different levels of comprehension depicting “how well a user understands (and, consequently, is able to use) an ambient display.” Shami et al. evaluated peripheral displays according to their context of use [12]. As part of the scenario-driven method the authors proposed a questionnaire that enables users to reflect on the display design informed by the main display characteristics. The resulting questionnaire contains questions grouped within the categories noticeability, comprehension, relevance, division of attention, and engagement. Finally Matthews et al. analysed the definition, design, and evaluation of peripheral displays using activity theory [6] and came up with a set of evaluation criteria including appeal, learnability, awareness, and distraction.

Based on these existing evaluation guidelines and inspired by concepts coming from the social and behavioural sciences intrusive and non-intrusive evaluation styles for ambient displays have been proposed [13] and illustrated with case studies [14]. Using the intrusive style the user is “consciously disrupted by the evaluation” method, thus providing reliable quantitative measures with lower validity due to higher cognitive load. The non-intrusive style abstains from disrupting the user and “focuses on the actual use in a general environment (in situ) over a long period.” In doing so the cognitive load is lowered and the gathered results have a higher validity, but the mostly qualitative measures are also less reliable. Nevertheless, non-intrusive evaluation techniques present a promising approach for ambient display evaluation. In combination with established methods to measure knowledge transfer the user attention towards ambient displays can be understood and evaluated more holistically. Thus facilitating an attention-aware ambient display design and a more reliable measurement of their impact.

**Purpose**

In earlier work we presented a conceptual framework to use ambient displays in a learning context. The framework is based on informational, interactional, and instructional characteristics and consists of parts dedicated to user and context data acquisition, channelling of information, and delivery of contextualised information framed in a learning process as described by Börner et al. [1]. The framework incorporates the four design dimensions of ambient systems introduced by Pousman & Stasko [11], namely information capacity, notification level, representational fidelity, and aesthetic emphasis.

This study pursues our research with a focus on the interaction between users and ambient displays and its influencing factors. We aimed to examine the user attention towards the display, how this is related to the interactions taking place, and what user reported information can (additionally) tell us about it. We see user attention as a critical factor to involve the user in a more intense interaction with the ambient display and thus as one fundamental requirement to eventually provoke learning. Consequently, our main purpose was to examine the user attention towards ambient displays as well as the influence of different display designs. The results should support the hypothesis that variations in the display’s design affect the user attention towards the display.

**FORMATIVE DESIGN STUDY**

In the context of this study the developed ambient display prototypes were varied on the design dimensions notification level and representational fidelity as defined by Pousman & Stasko [11]. The prototypes responded mainly to the first and the second phase of the introduced audience funnel framework [10], namely passing by as well as viewing and reacting. Following the suggestion of Wood et al. [16] we incorporated knowledge transfer on a per user basis as another evidence assuming that a better knowledge transfer is another indicator for an effective attention design. Therefore the prototypes provided factual knowledge to the users. Based on factors and metrics defined and discussed in related work (e.g. [4,6]) the criteria of interest were noticeability, disruption, comprehension, appeal, and relevance. The assumption was that the evaluated criteria have a direct influence on the knowledge transferred.

**Method**

For the experimental variation two independent variables were defined, i.e. the representational fidelity as well as the level of notification of the ambient displays, each variable could take one of two values. This resulted in four different treatments combining the two variables, i.e. ambient display prototype with either 1) change blind/indexical, 2) change blind/symbolic, 3) interruptive/indexical, or 4) interruptive/symbolic notification and representation.

As dependent variables the attention towards the display as well as the components noticeability, disruption, comprehension, appeal, and relevance have been measured. The user attention was measured directly during the treatment utilising a non-intrusive evaluation technique. The components noticeability, disruption, comprehension, appeal, and relevance were measured with post-test questionnaires after the treatment. Each component comprised a rating scale item.

The study was designed in the context of an institutional energy conservation project with the goal to promote conservation activities and reduce the overall energy consumption of the workplaces located at the main campus of a university. The study involved the four distinct experimental treatments with a post-test after the experiment. The experiment was performed for four consecutive weeks.
Participants
A total of 563 university employees were asked to participate in the study. 101 employees responded to the post-test. The prototypes were deployed in the entrance areas of the four main buildings of the university site. Only employees working in one of these main buildings were considered as participants of the experiment, yielding 94 post-test respondents for analysis. These participants were divided into groups depending on the building they are working in. The 94 participants (37 females and 57 males) were aged between 26 and 65 and had been working for the university for between two and 26 years. Because of the assignment procedure the study implemented a quasi-experimental research design where participants are not distributed to groups at random [2].

Based on the experimental variation 35 participants were exposed to the prototype with the change blind notification and indexical representation treatment ($N_1$=35), 12 to the prototype with the change blind notification and symbolic representation treatment ($N_2$=12), 12 to the prototype with the interruptive notification and indexical representation treatment ($N_3$=12), and 35 to the prototype with the interruptive notification and symbolic representation treatment ($N_4$=35).

Apparatus
For the experiment four prototypes were deployed in the four chosen campus buildings during the workdays and working hours of the university. All four settings were comparable in terms of lighting and other environmental conditions. Corresponding to the main characteristic of ambient displays [1], i.e. delivering information out of the periphery of attention, while being able to move between the periphery and the focus of attention, prototypes were developed and used to emulate this functionality. Each prototype consisted of a 20-inch laptop screen with built-in speakers and webcam but without attached keyboard or mouse. The speakers were used to send out audio notifications, while the webcam was used to enhance the sensorial functionality of the laptop. The prototypes presented pre-compiled slides showing three types of information, divided into three parts: a) depicting information regarding energy consumption, b) generic saving tips, c) and the overall conservation potential.

The prototype variation on notification level was implemented using a custom-built movement/attention sensor to trigger the notification as well as the built-in speakers to play back an audio notification. For the interruptive treatments one audio notification was played when the sensor detected movement in front of the display and another one when the sensor detected that someone looked at the display. For change blind treatments any notification was omitted. The variation on representational fidelity was implemented as two distinct means of information presentation. For the indexical representation raw data facts were used to communicate consumption information, saving tips, and conservation potentials. In contrast, topic-related icons were used for the symbolic representation of the data, e.g. light bulb icons representing 5W each.

Instruments
Two distinct instruments were used in the context of the study. The user attention towards the display prototypes was measured using custom-built movement/attention sensor. The sensor measured (during the experiment) any movement in front of the display as well as the number of users looking at and thus attending the display. This concept is based on the work of Shen et al. [13,14] but differs in its technical implementation. The sensor was solely built using the Processing’s development environment and the open source computer vision library OpenCV for Processing. Using the included standard image processing algorithms the sensor calculated differences between consecutive webcam images to detect movement and used the OpenCV face detection (standard Haar classifier cascade) to derive the number of users looking at the display. Thereby movement was only detected if the calculated difference between two images exceeded a predetermined (tested) threshold value. Besides triggering possible notifications (depending on the prototypical variation) these activities were stored in separate log files. Each entry in the log files lists an activity timestamp as well as information about the experimental treatment of the prototype. During the experiment the sensor’s validity in recognising movement and detecting faces has been checked randomly by reviewing the webcam images whenever a face was detected. To avoid privacy issues neither the images used for processing nor the detected users were stored. Two parameters could be analysed from the gathered log data of the custom-built movement/attention sensor: number of users passing by the display and number of users looking at the display.

On a user basis both parameters indicated in which phase (i.e. passing-by or viewing and reacting) of the audience funnel framework [10] the user was in. The actual user attention towards the display was then defined as the ratio between the number of users looking at the display and the number of users passing by the display. Consequently, this ratio depicts the effectiveness of the display to facilitate the transition from the passing-by to the viewing and reacting phase: the higher the ratio the more effective the display’s design in this context.

In addition to that a post-test questionnaire measured the individual noticeability, disruption, comprehension, appeal, relevance, and knowledge transfer to provide further evidence for the actual user attention towards the display. Thereby the transferred knowledge indicates to which degree a user advanced to the third audience funnel phase

1 http://processing.org
(i.e. subtle interaction): the higher the score the more effective the display’s design to facilitate the transition from the underlying phases. Inspired by the reviewed literature the questionnaire has been constructed specifically for this study and contained beside the measured components also demographic questions, such as year of birth, gender, period of employment, as well as workplace location. The questionnaire measured the single components with distinct items, e.g. “Did you consider the information display as disrupting?”

The data was collected quantitatively. The used 7-point scaled items provided an open range of choices from 1 (not at all) to 7 (completely) to express the participants’ agreement regarding a statement. To assess the actual knowledge transfer the questionnaire also contained the following multiple-choice question: “Can you specify what kind of information has been presented on the display?” The possible responses included information regarding energy consumption, generic saving tips, and overall conservation potential. For a fully recognised knowledge transfer all response options had to be checked.

Per day on average 969 users passed by the four display prototypes and 87 users looked at them. The highest number of users passing by the displays as well as the highest attention towards them was observed during noon. Although more users passing by were observed in the first half of the day, the average attention ratio is slightly higher in the afternoon hours compared to the morning.

To derive the influence of the independent variables (representational fidelity and notification level) on the attention ratios the experimental groups were compared. All groups started with a high ratio in the first week. The highest ratio (22%) was observed for the group with symbolic representation and change blind notification, while the group with indexical representation and change blind notification had the lowest ratio (9%). In the second week the ratio dropped almost to the same value for all groups (around 5%). In the third week the ratio then increased for the group with symbolic representation and change blind notification (21%) as well as the group with indexical representation and interruptive notification (20%), while remained consistent for the other two groups. This development also continued in the last week of the experimental study with a slightly decreasing ratio for the groups with the indexical representations. Based on the observations no clear statement can be made on the influence of the independent variables on the attention ratio.

Based on the assumption that the observed novelty effect levels off in the course of the experimental study [3] and thus the attention ratio towards the displays reaches a consistent level, the ratios observed in the last week of the experimental study were taken for further analysis. Table 1 lists the values.

Table 1. Averaged attention ratios observed in the last week of the experimental study for each treatment and assigned group.

<table>
<thead>
<tr>
<th>Design dimensions</th>
<th>Representational Fidelity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Indexical</td>
</tr>
<tr>
<td>Notification Level</td>
<td>Change blind</td>
</tr>
<tr>
<td></td>
<td>A&lt;sub&gt;1&lt;/sub&gt; = 1%</td>
</tr>
<tr>
<td></td>
<td>Interruptive</td>
</tr>
<tr>
<td></td>
<td>A&lt;sub&gt;3&lt;/sub&gt; = 17%</td>
</tr>
</tbody>
</table>

The differences between the ratios when keeping one independent variable constant were tested on statistical significance using two-proportion z-tests, resulting in four distinct one-tailed tests for each manifestation of the independent variables. This approach is appropriate because the sampling method was simple random sampling, the samples were independent, each population was at least 10 times larger than its sample, and each sample included at least 10 successes and 10 failures. The result shows that for displays with an interruptive notification level the attention...
A number of regression analyses were used to examine if change blind notification and disruptive representation (Group 4, with the interruptive notification and symbolic group) medians for disruption suggests that the prototype significant difference on disruption between Group 2 and 4 noticeable (Group 3, according to the participants the prototype with the 0.41). Inspecting the groups’ medians shows that between Group 3 and 4 (3). The test for disruption revealed a significant effect of the ordinal scale and thus do not meet t-tests. The test revealed a significant effect of the treatments on notice (H(3) = 7.88, p = .049) and disruption (H(3) = 8.48, p = .037). Mann-Whitney tests were conducted to follow-up the findings, controlling for Type I error across tests by using the Bonferroni approach. The tests for notice revealed a significant difference on notice between Group 3 and 4 (U = 112, z = −2.79, p = .005, r = −.41). Inspecting the groups’ medians shows that according to the participants the prototype with the disruptive notification and indexical representation (Group 3, N3=12) was the least noticeable (Median = 5.5), while the prototype with the interruptive notification and symbolic representation (Group 4, N4=35) was the most noticeable (Median = 7). The test for disruption revealed a significant difference on disruption between Group 2 and 4 (U = 93.5, z = −2.89, p = .004, r = −.42). Inspecting the groups’ medians for disruption suggests that the prototype with the interruptive notification and symbolic representation (Group 4, N4=35) was considered most disruptive (Median = 4), while the prototype with the change blind notification and symbolic representation treatment (Group 2, N2=12) was considered least disruptive (Median = 1.5).

A number of regression analyses were used to examine if single or multiple factors influenced each other. The results of the analyses indicated that:

• The rated appeal of the information visualisation explained 32% of the variance in perceived usefulness and relevance of the presented information (R² = .32, F(1,92) = 43.14, p < .001). Appeal significantly predicted relevance with β = .57, t(92) = 6.57, p < .001.

• The noticeability of the display explained 27% of the variance in comprehension of the information given (R² = .27, F(1,92) = 33.45, p < .001). Notice significantly predicted comprehension with β = .52, t(92) = 5.78, p < .001.

• The rated appeal of the information visualisation explained 22% of the variance in the perceived disruptiveness of the display (R² = .22, F(1,92) = 25.25, p < .001). Appeal significantly predicted disruption with β = −.46, t(92) = −5.03, p < .001.

The actual knowledge transferred was assessed with a multiple-choice question asking for the different kind of information presented. The responses were then scored depending on the number of options checked with 1 for one option checked, 2 for two options checked, 3 for three options checked, and 0 for none or “other” option checked. On average the participants scored moderately across all display prototypes (Median = 2) with the best score frequency for the group exposed to the prototype with the interruptive notification and symbolic representation treatment (Group 4, N4=35). A Kruskal-Wallis test showed that the differences between the treatment groups are not significant. Furthermore a logistic regression analysis was used to examine if single or multiple factors influenced the knowledge transferred. The results indicate that comprehension of the information given as well as the perceived usefulness and relevance of the presented information are significant predictors for the knowledge score. Comprehension significantly predicted whether participants scored 0 or 1, 2, and 3 respectively. Relevance significantly predicted whether participants scored 0 or 1 on knowledge.

DISCUSSION
The results of the presented formative design study show a high degree of user interest in the displays over time. The highest attention rate has been measured during the first days of the study, while it peaked again in the middle and at the end of the study. The novelty effect can be accounted for some of the observed variance in the data especially in the beginning of the study. In general the number of users passing by is evenly distributed, while certain days are clearly busier than others. The user interest did not really stabilise in the course of the study. In line with the findings of Shen et al. [14] this calls for an even longer evaluation period. Although more people passed-by the displays in the morning hours, the highest attention ratio has been measured in the second half of the day with a peak around noon when users usually leave the building for a lunch break. Eventually the effectiveness of ambient displays depends on these surrounding conditions, e.g. differences in the frequency of users passing-by during the day. Existing
daily routines need to be considered and the display design should be adapted accordingly. A stronger contextualisation of the information presented might be one solution.

The results are inconclusive regarding the initial hypothesis on the effectiveness of the chosen representational fidelity and the level of notification. However the results suggest that the chances are higher to get the user attention when designing ambient displays with easy to grasp information and a sensible but not demanding level of notification. The challenge is to find the right balance. The attention ratio towards the ambient display is higher when combining interruptive notification with indexical representation or symbolic representation with change blind notification. The reasons for this are manifold, but from a user perspective it’s obvious that when something draws immediate attention users prefer a fast and direct access to the conveyed information. On the other hand it’s also obvious that a contrasting design seduces users by responding to their curiosity regarding the surrounding triggered by their peripheral perception capabilities. The non-intrusive provision of tangible information suits this purpose best.

Besides looking at the quantitative attention data the study tried to support it’s claims with additional qualitative data. Supporting the conclusions drawn from the actual attention measurement, the reported results are inconclusive regarding the reported disruptiveness and comprehension of the displays. Participants take more notice of an interruptive display presenting symbolic information. At the same time they also felt more disrupted by it. These factors influence each other, e.g. it can be argued that the more disruptive a display is perceived the more users (at least initially) take notice of it. The study results do not provide evidence for this claim but reveal other potential relationships, especially the high impact of an appealing information visualisation. On the one hand the presented information is perceived more useful and relevant, while at the same time the information display is considered less disruptive. The noticeability of the display improved the comprehension of the information presented.

Following the initial hypothesis this should also affect the knowledge transferred and thus provide another indicator regarding the user attention. Again the results are inconclusive regarding the effectiveness of the chosen representational fidelity and level of notification. Nevertheless they provide evidence for the importance of providing comprehensible and relevant information. Thereby the perceived usefulness and relevance of the presented information acts as a trigger for the knowledge transfer. Again the result calls for a contextualisation of the information presented. Once started the reached level of comprehension then facilitates the process further.

In sum the results are self-contradictory reflecting some of the tensions for an effective attention-aware ambient display design. The presented study has several limitations regarding the measurement and analysis of user attention towards ambient displays. The quantitative approach using sensor data is no reliable measurement of user attention. Single users cannot be identified and thus no statistical methods can be applied for the analysis. The data only presents a rough estimation of the actual user attention. The qualitative approach helps but does not solve this problem completely. The gathered questionnaire data is no valid measurement of user attention. Single users can be identified and thus statistical methods can be applied to analyse the data. Still the used questionnaire is no conclusive inventory of user attention. Nevertheless the combination of quantitative and qualitative measurement does provide a more holistic view on user attention.

**CONCLUSIONS**

The presented study focused on the interaction between ambient displays and users. The study combined non-intrusive evaluation techniques as a quantitative approach to measure user attention with qualitative measurement of user perception and comprehension. The results helped to understand and estimate the potential of the introduced ambient display prototypes for learning. The hypothesis that variations in the display design (i.e. on the introduced dimensions notification level and representational fidelity) affect the user attention towards the display sustainably could not be confirmed. Nevertheless, several guidelines for an effective attention-aware ambient display design can be derived.

The introduced prototypes were mapped to the first and the second phase of the audience funnel framework [10], namely passing-by, and viewing and reacting. To overcome the threshold between the phases the display needs to attract the users’ attention. To do so successful display designs need to be contextualised and should not go beyond ‘just the right’ level of obtrusiveness while providing glanceable information. A possible bias is the novelty effect that accounts for outstanding user attention shortly after deployment and then levels off quickly. To advance even to the third phase of the audience funnel framework (i.e. subtle interaction) and finally get the desired message through, users need to be intrigued and motivated. The intended transfer of knowledge is initiated when the presented information is both relevant and appealing. Finally comprehensibility facilitates this process.

Confirming McCrickard et al. [7] the main challenge for an effective design is the right balance of attention and usability. How to deal with this trade-off is mainly influenced by the intended use of the ambient display. Several contextual factors have an impact on the design. This study presents a first attempt to identify and relate some of these determining factors from a user perspective especially in a learning context. The gathered insights can also inform future research and developments beyond that with a focus on a coherent attention-aware and contextual design of ambient displays.
REFERENCES


