

Heuristic Evaluation of Ambient Displays

Jennifer Mankoff¹, Anind K. Dey², Gary Hsieh¹, Julie Kientz³, Scott Lederer¹, Morgan Ames¹

¹EECS Department, UC
Berkeley
Berkeley, CA, USA

²Intel Research
Berkeley, CA, USA
ambient@cs.berkeley.edu

³University of Toledo
Toledo, OH, USA

ABSTRACT

We present a technique for evaluating the usability and effectiveness of ambient displays. Ambient displays are abstract and aesthetic peripheral displays portraying non-critical information on the periphery of a user's attention. Although many innovative displays have been published, little existing work has focused on their evaluation, in part because evaluation of ambient displays is difficult and costly. We adapted a low-cost evaluation technique, heuristic evaluation, for use with ambient displays. With the help of ambient display designers, we defined a modified set of heuristics. We compared the performance of Nielsen's heuristics and our heuristics on two ambient displays. Evaluators using our heuristics found more, severe problems than evaluators using Nielsen's heuristics. Additionally, when using our heuristics, 3-5 evaluators were able to identify 40-60% of known usability issues. This implies that heuristic evaluation is an effective technique for identifying usability issues with ambient displays.

KEYWORDS: Ambient displays, heuristic evaluation, discount evaluation

INTRODUCTION

Ambient displays are aesthetically pleasing displays of information which sit on the periphery of a user's attention. They generally support monitoring of non-critical information. They represent a burgeoning sub-area of ubiquitous computing, and examples of these displays have been recently published at CHI, UIST, Ubi-comp, DIS, as well as being displayed in museum and art venues. Although many interesting displays have been created since Weiser first described the dangling string [22], little is known about what makes one ambient display more effective than another, or how to evaluate these displays.

Ambient displays have the ambitious goal of presenting information without distracting or burdening the user. This goal is difficult to design for and difficult to define in measurable terms. Because of this, evaluation of am-

bient displays is difficult, and has been limited in the past to formative ethnographies [14], and iterative, "living laboratories" [12] in which applications are evolved over time as they are used by their creators. However, both of these techniques are costly and time consuming. As a result, most ambient displays have not been evaluated at all. Without evaluation, it is hard to determine which displays are effective and why they are effective. Without this information, it is difficult to improve on existing work.

Inexpensive, or *discount*, formative techniques could provide guidance at the early stages of design without requiring the intense time commitment of an ethnography. However, existing formative techniques, such as GOMS and heuristic evaluation are focused on systems with clearly defined tasks, and goals such as productivity and efficiency. For example, one of Nielsen's heuristics calls for documentation "focused on the user's task" [19]. In contrast, ambient displays are at the periphery of the user's attention and main task, and have as primary goals relevancy of information, aesthetics, and ease of monitoring. While there is some overlap here, we believe the differences are significant enough to bring into question the applicability of existing evaluation techniques to the domain of ambient displays.

This work focuses on adapting heuristic evaluation because it is a widely used, inexpensive, formative evaluation technique that we believe could be a useful tool for ambient display designers. Additionally, Jakob Nielsen, the creator of heuristic evaluation, wrote that "the difficulties in user testing some next-generation interfaces may mean greater reliance on the heuristic evaluation method" [16]. Similar to Baker *et al.*, who applied heuristic evaluation to the domain of CSCW, we believe that the way to adapt heuristic evaluation is through modification of the heuristics [1].

We start by giving an overview of evaluations of previous ambient displays. Next, we describe two ambient displays we have built. Although we provide evidence that these are usable displays, the focus of this paper is not their usability, but rather the applicability of heuristic evaluation to ambient displays. We discuss how heuristic evaluation is currently used, and present a set of heuristics created with the help of expert ambient display designers to address the idiosyncratic requirements of ambient displays. We designed an experiment

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2003, April 5-10, 2003, Ft. Lauderdale, Florida, USA.

Copyright 2003 ACM 1-58113-630-7/03/0004...\$5.00.

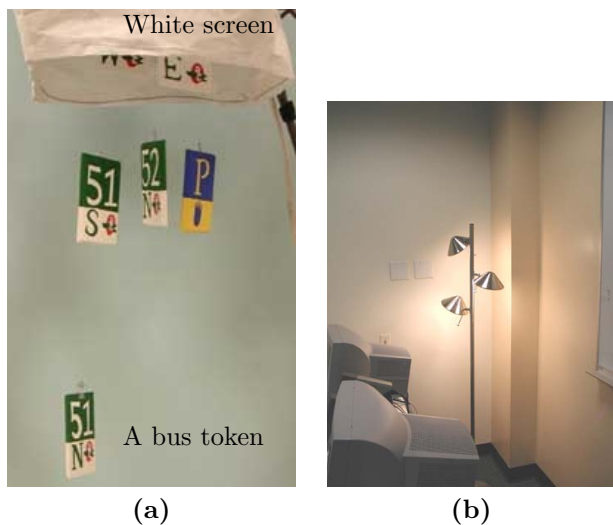


Figure 1: (a) BusMobile. (b) Daylight Display.

to compare the effectiveness of Nielsen’s heuristics with our set. After discussing the experimental design and results, we present conclusions and future work.

RELATED WORK

The majority of ambient displays that have been published have not been evaluated, (*e.g.* [22, 7, 6, 9]), or report only short evaluations with few details or informal feedback, focusing instead on technology and design innovation [10, 3, 13]. Some displays have been shown in museum settings where they were used by hundreds of users but never tracked in detail [2, 8, 23].

Some of the most extensive evaluation effort has been focused on determining exactly what users might want, using techniques such as surveys, interviews, and “Wizard of Oz” prototyping [14]. Additionally, researchers investigating displays that combine ambient awareness with notification, or alarms, have conducted some formative analysis [4, 11]. In particular, Chewar and McCrickard have discussed work in progress on modifying a formative in-use study methodology to apply to notification displays [4]. The same group is also investigating perceptual issues relevant to the design of ambient and notification displays [20].

The most common “finding” of exhibits and empirical studies of ambient displays is that users are interested in and excited by innovations in ambient displays. In contrast, our contribution is a better understanding of how to apply a discount evaluation technique, heuristic evaluation, to ambient displays.

TWO DISPLAYS

This section describes the two ambient displays that we evaluated with heuristic evaluation. Both displays are shown in Figure 1. We created these displays using user-centered design techniques as follows: First, we surveyed undergraduates using the windowless labs of our computer science building in order to determine what in-

formation would be of interest to them. We chose the highest ranked information sources that included dynamic change (and thus were amenable to an ambient display approach). We then built the displays and deployed them in the aforementioned labs for a two week period.

The first display was designed to provide information about popular bus lines. Specifically, it indicates how close each commonly used bus (as reported by the students) is to the nearest bus stop. Closeness is determined from the published bus schedules and current time of day. As a bus approaches a bus stop, the token representing that bus moves up toward the top of the display (see Figure 1(a), which shows that Bus 51N is further from the bus stop than Bus 51S). Tokens remain out of sight under the white screen until a bus is less than 25 minutes away, and then move down to the lowest possible depth in order to start their “approach.”

The second display was designed to provide information about whether it is dusky, light, or dark outside. This was of interest to students for safety reasons (*i.e.* some students are uncomfortable walking home alone after dark) and also as an approximate indicator of time of day. The display flickers when evening “begins” (this was determined by observing how long before sunset it begins to get dark out), and then slowly fades from light (bright) to dark (dim). It does the opposite at dawn.

Throughout the rest of this document, we refer to the first display as the “busMobile” and the second as the “daylight display”. Qualitative feedback indicated that students preferred the busMobile to the daylight display. For example, one student wrote “bus mobile ->ultra cool. makes life easier” about the busMobile. Written comments criticized the daylight display for being too bright, or indicated that a respondent had not noticed it or thought it was broken.

We surveyed a total of 60 students during the study period. There was a strong correlation (Spearman’s $\rho=.618$, $p<.01$) between interest in the daylight information (based on a Likert scale response of 4 or 5 out of 5) and those who reported finding the daylight display useful. There was a very strong correlation ($\rho=.808$, $p<.01$) between interest in the bus schedules and the usefulness of the busMobile, while there was a moderate correlation ($\rho=.595$, $p<.01$) between interest in the bus schedules and respondents who missed it after it was removed.

In summary, we used a user-centered design process to build two ambient displays and gain feedback on their usefulness and some of the problems with their design. Since little is known about the usability of other published ambient displays, we felt it was important to build *and test* the displays we would give to our heuristic evaluators in this work in order to verify that they were at

least minimally appropriate to the task for which they were designed. Our evaluation showed that these displays are useful to users interested in the information they display. Note that the rest of this paper *is not* concerned directly with the usability of these displays in particular. This paper is focused on adapting heuristic evaluation to ambient displays. The displays reported in this section are a vehicle for comparing the effectiveness of Nielsen’s heuristics against our own.

HEURISTICS: ORIGINS, USE, AND MODIFICATIONS

We believe heuristic evaluation is important to ambient display designers because of its potential to provide quick, inexpensive feedback about the possible issues with a specific display [18]. Note that heuristic evaluation is most effective when performed by trained usability evaluators and does not help to identify which of these issues are real problems and will have a measurable impact on usability [5].

Heuristic evaluation involves recruiting evaluators, who may be novices, to critique an interface (usually represented with pictures and a textual description). Evaluators look for problems in an interface’s compliance with heuristics that encode important usability guidelines. Nielsen found that 3-5 novice evaluators find 40-60% of known issues when applying heuristic evaluation [18]. The canonical list of heuristics as defined by Nielsen is [15]: (1) Visibility of system status; (2) Match between system and the real world; (3) User control and freedom; (4) Consistency and standards; (5) Error prevention; (6) Recognition rather than recall; (7) Flexibility and efficiency of use; (8) Aesthetic and minimalist design; (9) Help users recognize, diagnose and recover from errors; (10) Help and documentation.

Heuristic evaluation, because of its informal nature and low cost, was rated as one of the top techniques currently in use in a survey of usability practitioners [21]. Among discount evaluation techniques, only “informal expert review” ranked higher than heuristic evaluation. However, heuristic evaluation was considered to be more effective than informal expert reviews by the survey participants. Also, it was believed to be low cost and fast, and easy to apply. Cost, and speed were both highly ranked factors in choosing an evaluation methodology.

The major difference between evaluating ambient displays and evaluating traditional displays comes from the way users interact with the interface. Ambient display users are passive in obtaining information from a display. Users do not *use* the displays as they would use computers; they *perceive* the displays. Consequently, some of Nielsen’s original heuristics (those associated with interactivity and productivity) do not apply to ambient displays. Furthermore, since ambient displays are passive displays, more responsibility lies with the designers to ensure that the displays are conveying a sufficient

amount of information, and that the information is being conveyed unobtrusively. We were particularly concerned with how Nielsen’s heuristics were defined, not only how they were named, as the definitions made assumptions about the goals of the interface being evaluated. For example, two heuristics explicitly refer to a “dialog” with which the user interacts, something that is central to conventional interfaces, but rare in ambient displays. Our conclusion was that the methodology of heuristic evaluation could be applied to ambient displays if the heuristics used were modified.

In summary, we chose heuristic evaluation because it is one of the top evaluation techniques in use. However, it does not address issues that are important to the design of ambient displays. Nielsen’s heuristics are not appropriate for ambient displays, because they assume interactive, productivity-oriented systems, while ambient displays may not be interactive, and support monitoring rather than efficiency. We hypothesized that a modified set of heuristics could find more usability issues if it took these aspects into account.

In this section, we present the results of a survey of experts in the design and analysis of user interfaces and experts in the design of ambient displays. We asked them to rate the modified set of heuristics we had created and add any they thought were missing. We revised our heuristics based on these results, and then asked a group of experienced user interface evaluators and ambient display designers to evaluate the two displays described above using both sets of heuristics. That evaluation is described in the next Section.

Research method

Our goal in this study was to obtain feedback on a set of heuristics that was based on Nielsen’s, but modified to be more applicable to ambient displays. Nielsen’s heuristics each consist of a title and definition. We eliminated non-applicable heuristics (numbers 3,4,5,7,9,10), based on two independent reviews of the heuristics by separate researchers in our group. Applicability was based on how well the heuristics met the primary goals of ambient displays. We modified some heuristic titles and definitions in the remaining Nielsen’s heuristics in terms more appropriate to ambient displays, and added 5 heuristics specific to ambient displays based on the two reviews and a group brainstorming session.

We further modified these heuristics after running a pilot survey with local ambient display designers. Table 1 shows the modified heuristics (left column). We then surveyed outside experts in order to produce the final heuristics shown in the right column of Table 1.

Four ambient display designers (domain experts), two usability experts, and one visual designer participated in our survey. The median years of experience with heuristic evaluation, and with ambient displays were both 4,

Modified heuristics:	Rel.	Final Heuristics:
Useful and relevant information The information should be useful and relevant to the users in the intended setting.	4.5	Useful and relevant information The information should be useful and relevant to the users in the intended setting.
“Peripherality” of display Display should be unobtrusive and remain so unless it requires the users’ attention.	4.8	“Peripherality” of display The display should be unobtrusive and remain so unless it requires the user’s attention. User should be able to easily monitor the display.
Match between aesthetics of ambient display and environment One should notice an ambient display because of a change in the data it is presenting and not because its design clashes with its environment.	4.6	Match between design of ambient display and environments One should notice an ambient display because of a change in the data it is presenting and not because its design clashes with its environment.
Sufficient information design The display should be designed to convey “just enough” information. Too much information cramps the display, and too little will make the display not as useful as it could be.	4.4	Sufficient information design The display should be designed to convey “just enough” information. Too much information cramps the display, and too little makes the display less useful.
Recognition rather than recall Ambient displays are meant to reduce cognitive load, which is impaired when users must remember what states or changes in the display mean. The display should be intuitive.	4.7	Consistent and intuitive mapping Ambient displays should add minimal cognitive load. Cognitive load may be higher when users must remember what states or changes in the display mean. The display should be intuitive.
Easy transition to more in-depth information The display should make it easy and quick for users to find out more detailed information, if the display offers multi-leveled information.	4.4	Easy transition to more in-depth information If the display offers multi-leveled information, the display should make it easy and quick for users to find out more detailed information.
Visibility of state An ambient display should make the states of the system more visible and accessible. The transition from one state to another should be easily noticeable.	4.4	Visibility of state An ambient display should make the states of the system noticeable. The transition from one state to another should be easily perceptible.
	N/A	Aesthetic and Pleasing Design The display should be pleasing when it is placed in the intended setting.
Display better than its replacement The display should be better than what it replaces or supplements.	4.2	

Table 1: A comparison of the modified heuristics used in the survey (left), and the final heuristics derived from the survey results (right). The middle column shows the average relevance assigned to each heuristic by respondents (5 being most relevant).

with averages of 5.4 and 3.2, respectively. Self-rated experience levels for heuristic evaluation and ambient displays both averaged 3 on a 5 point Likert scale (5 being highest).

Participants were given the following definition for ambient displays:

An ambient display normally communicates on the periphery of human perception, requiring minimal attention and cognitive load. Perceptual bandwidth is minimized; users get the gist of the state of the data source through a quick glance, aural refocus, or gestalt background ambience. A common usage for ambient displays is to support the awareness of data sources.

Ambient displays are not limited to only visual displays. A slightly opened door is an example of an ambient display, and the information that one obtains through the peripheral senses could be visual, audio, and/or olfactory.

Each participant was asked to provide a relevance rating on a scale of 1 to 5 (5 being highest) and comments about each of the heuristics shown on the left in Table 1. Participants were also encouraged to suggest additional heuristics at the end.

Results

The average relevance ratings were very high for each heuristic, with none lower than 4.2. The comments section of the survey provided the most useful data for revising the heuristics. For example, one participant commented that if an ambient display were to be effective, it should be calming, and two commented that ambient displays should not be interruptive. Another respondent went further by suggesting that ambient displays should increase a user’s enjoyment or engagement. These comments caused us to add a heuristic not present in our original list, “Aesthetic and Pleasing Design.” We also

removed the lowest rated heuristic “Display better than its replacement” because of repeated criticism of the word “better” and the lack of specificity in the heuristic’s definition. Our other changes were more subtle, such as changing the word “accessible” to “noticeable” in our definition of “Visibility of state”, and renaming “Recognition rather than recall” to “Consistent and intuitive mapping” so that it would better match the definition we provided in the survey.

COMPARING HEURISTICS

After our modification of the heuristics for ambient displays, our next step was to apply the modified heuristics to actual ambient displays using a formal heuristic evaluation methodology. The purpose of this study was to compare the effectiveness of our heuristics to Nielsen’s heuristics. We tried to replicate Nielsen’s original work in our study design [18]. We will refer to evaluations done with Nielsen’s heuristics and the modified heuristics as the “Nielsen” and “ambient” conditions, respectively. Our hypotheses were:

1. The number of issues found in the ambient condition will be greater and the issues will be more severe than those found in the Nielsen condition.
2. The percentage of known issues found in the ambient condition will be higher than the percentage of known issues found in the Nielsen condition.
3. The ambient heuristics will be more useful to evaluators than Nielsen’s heuristics. A heuristic that finds many, severe problems is more useful than a heuristic that finds fewer problems with lower severity.

Method

Participants. Sixteen participants were recruited, with a median of 5 years of evaluation experience. We split them into two groups of 8 people. We balanced ambient display expertise between the groups, since there was a high variance with 10 participants having less than 1 year of ambient display experience, and the others having 2 or more.

Five users is typically cited as sufficient to find 80% of issues when conducting a heuristic evaluation [17]. However, Woolrych and Cockton [24] criticize the claim that five users is sufficient on the basis that problems may not be evenly distributed by severity, and users may not consistently find 30% of problems (an assumption of Nielsen and Landauer’s original claim [17]). These criticisms are directed not at the technique itself, simply at the number of users necessary to conduct an effective evaluation. We included 8 participants in each condition to address these criticisms.

Procedure. We used a between-subjects design in which group one was asked to evaluate the busMobile and the daylight display using Nielsen’s original heuristics [15] (the Nielsen condition), and group two was asked to

Severity	Bus	Light	%N	%a	%n
1	4	4	63%	63%	0%
2	15	4	63%	58%	21%
3	7	7	50%	93%	0%
4	8	7	47%	80%	13%
5	0	2	0%	100%	0%

Table 2: The number of known issues at each severity rating for each display. The fourth and fifth columns show the percentage of these issues found with Nielsen’s (%N) and with the ambient (%a) heuristic sets. Some problems were not found with either heuristic set (%n)

evaluate the same displays using the modified heuristics (the ambient condition). The order in which the displays should be evaluated was randomly assigned.

Task. Each participant completed the evaluation individually, and was allowed to spend as much time as she chose on each evaluation. Participants were sent an email containing the heuristics they should use. The email also specified the evaluation order for the displays. The “description” of each display was a web site which participants were directed to look at. This web site was identical for all participants in both subject groups.

An introductory web page described the purpose of the study – to “adapt heuristic evaluation to ambient displays” – and defined ambient displays and heuristic evaluation. Each display’s web page included a description of the ambient display, the setting in which it was used, and images and a written description of how it changed as data changed. Participants were directed to read these web pages, and then create a list of issues. Each issue was to include the heuristic or heuristics that it related to, and a severity rating.

Separately, following Nielsen’s methodology [18], we conducted an informal expert review locally in order to generate a master list of issues and their severities. *Severe issues* (ratings of 4-5 on a 5 point scale) were issues we felt would substantively discourage or impede a user’s ability to use the displays. *Minor issues* (ratings of 1-2) were deemed to represent irritations but not significant barriers to use. *Nonissues* (misunderstandings about how some aspect of the display worked, caused by the limited information given on the web pages) were given a rating of 0.

RESULTS

A total of 26 issues are known for the daylight display, 24 of which were found in the heuristic evaluation. Thirty-nine issues are known for the bus display, 35 of which were found in the heuristic evaluation. Two of the missed issues were severe, while four were not severe. The order in which displays were evaluated had no significant effect on the severity or number of issues found. Table 2 shows the number of known issues at each severity rating for each display.

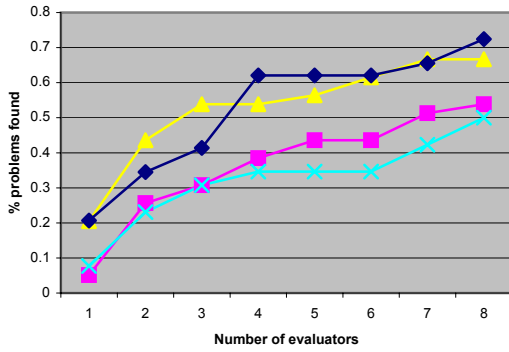


Figure 2: A comparison of the percentage of issues found by evaluators of each display using each set of heuristics. (Xs) daylight display, Nielsen's heuristics (Squares) busMobile, Nielsen's heuristics (Triangles) bus display, ambient heuristics (Diamonds) daylight display, ambient heuristics

Hypothesis 1: The number of issues found in the ambient condition will be greater and the issues will be more severe than those found in the Nielsen condition. We tested this hypothesis using a univariate analysis of variance test, comparing the average number of cases and average severity of each issue found with each heuristic set and display (busMobile and daylight display) as factors. By cases, we mean that if the same issue was found multiple times by different evaluators, we count each finding. There was no statistically significant difference for any factor or combination of factors in the average number or severity of issues found by evaluators.

Hypothesis 2: The percentage of known issues found in the ambient condition will be higher than the percentage of known issues found in the Nielsen condition. Figure 2 shows the increase in percentage of known issues found as the number of evaluators increases. It shows the percentage of issues found by evaluators in different conditions, for both displays. As shown in the figure, the percentage of issues found by all of the evaluators in the ambient condition, combined, is higher than the percentage of issues found by all of the evaluators in the Nielsen's condition, combined. Visual inspection shows that three to five novice evaluators find 40-60% of problems using the ambient heuristics, a result consistent with Nielsen's original work [18]. Graphs of two other randomly chosen orderings of evaluators showed similar results.

Figure 3 shows a breakdown of the issues found on a per-heuristic basis for the busMobile display. Note that the black bars (ambient condition) are often present without gray bars (Nielsen condition), indicating that an issue was found with the ambient heuristics but not Nielsen's heuristics. Additionally, when the bars are paired, the black bars tend to be taller. The far right of the barplot has a series of issues only found by Nielsen's heuristics. These were all minor issues or nonissues.

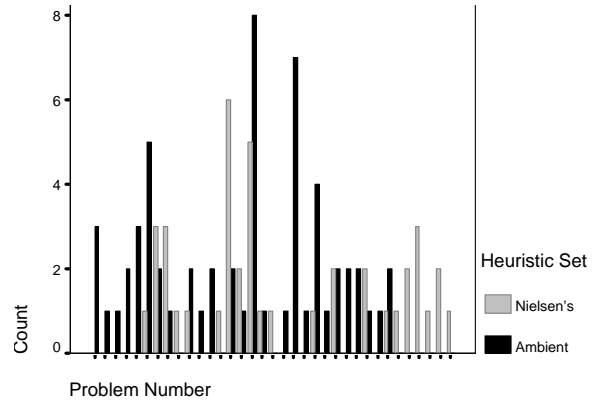


Figure 3: A comparison of the coverage of the ambient (black) and Nielsen's (gray) heuristics for the busMobile. The y axis shows the number of times an issue was found with a given heuristic set. The x axis indicates problem number. If no evaluator found an issue within one condition (heuristic set), no bar is shown for that condition.

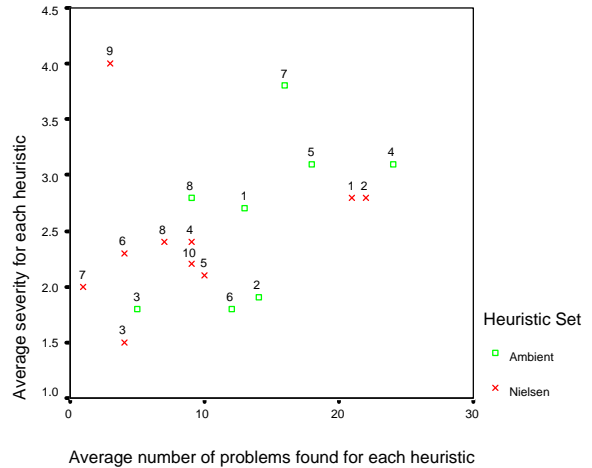


Figure 4: A plot of average issue severity against average number of issues found, for each heuristic.

Hypothesis 3: The ambient heuristics will be more useful to evaluators than Nielsen's heuristics. A heuristic that finds many, severe problems is more useful than a heuristic that finds fewer problems with lower severity. Figure 4 shows the correlation between the average number of issues found across all evaluators with each heuristic (including both the Nielsen and ambient heuristics) and the average severity of issues found with each heuristics. There is one outlier (a heuristic for which an average of 2 issues were found, with an average severity of 4), visible in the upper left. When that outlier was removed, a strong correlation proving our hypothesis was found (Pearson's $r = .657$, $p < .004$). Visual inspection shows that the ambient heuristics tend to be on the upper right of the plot. This is important because it means evaluators using the most useful ambient heuristics are finding a greater number of problems,

and problems of greater severity than evaluators using the most useful Nielsen's heuristics.

We also looked at the proportion of major (severity ≥ 4) issues and compared that to the proportion of minor (severity ≤ 2) issues found by each evaluator. On average, evaluators using the ambient heuristics found significantly more major issues (22%) than minor issues (12%) ($p < .05$). Evaluators using Nielsen's heuristics found only 11-13% of both major and minor issues.

DISCUSSION

The results of our evaluation indicate that the ambient heuristics were an improvement on Nielsen's heuristics for the domain of ambient displays. In order for heuristic evaluation to be effective, a small number of evaluators must be able to find a large percentage of known problems. A single evaluator using the ambient heuristics finds 22% of known major problems on average, and eight evaluators are sufficient to find about 70% of known problems. In contrast, using Nielsen's heuristics, a single evaluator will only find about 13% of major issues and eight evaluators find about 50% of known issues. The evidence suggests that the ambient heuristics will allow a small number of evaluators to identify most serious usability issues with a display.

Two of the eight evaluators in the Nielsen condition complained about the applicability of Nielsen's heuristics to ambient displays. One respondent gave us detailed comments on why they were inappropriate, while the other simply chose not to report any issues because he felt they were so irrelevant to the domain. Despite this, there was no statistically significant difference in the average number of issues found in each condition, across heuristics. We attribute this to the fact that many of Nielsen's heuristics are quite general, and evaluators were able to assign issues to those heuristics even when they were not an obvious fit. Additionally, inspection of the data shows that the ambient heuristics did not adequately address error conditions. For example, the problems shown on the far right of Figure 3, were not found by any evaluators using the ambient heuristics. Although these were minor problems, they highlighted the fact that the ambient heuristics do not include any mention of error states.

As further evidence that additional heuristics are needed, we found that to cover *all* of the known issues with a minimum number of heuristics, a combination of Nielsen's and ambient heuristics is needed. We repeatedly selected the heuristic responsible for the largest number of severe issues. After each selection, we removed all of the issues associated with the heuristic from the list of remaining issues. We repeated this until no issues remained. The resulting selected heuristics included all but one of the ambient heuristics, and half of Nielsen's heuristics. This is not surprising given that, the most

useful ambient heuristics tended to perform better than the most useful of Nielsen's heuristics. This is the final set of heuristics, in order of issue coverage. The derived ambient heuristics are in italics, the others are direct quotes from Nielsen [15]:

Sufficient information design The display should be designed to convey "just enough" information. Too much information cramps the display, and too little makes the display less useful.

Consistent and intuitive mapping Ambient displays should add minimal cognitive load. Cognitive load may be higher when users must remember what states or changes in the display mean. The display should be intuitive.

Match between system and real world The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

Visibility of state An ambient display should make the states of the system noticeable. The transition from one state to another should be easily perceptible.

Aesthetic and pleasing design The display should be pleasing when it is placed in the intended setting.

Useful and relevant information The information should be useful and relevant to the users in the intended setting.

Visibility of system status The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

User control and freedom Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

Easy transition to more in-depth information If the display offers multi-leveled information, the display should make it easy and quick for users to find out more detailed information.

"Peripherality" of display The display should be unobtrusive and remain so unless it requires the user's attention. User should be able to easily monitor the display.

Error prevention Even better than good error messages is a careful design which prevents a problem from occurring in the first place.

Flexibility and efficiency of use Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

CONCLUSIONS AND FUTURE WORK

We have presented the iterative design of a set of heuristics applicable to ambient displays. We have shown these heuristics to be an improvement on Nielsen's heuristics for the domain of ambient displays, particularly in terms of the proportion of major problems found by individual evaluators, and the total percentage of known problems found by groups of evaluators. As an additional benefit, we have demonstrated a process for creating heuristics applicable to a new domain, building on the work of Baker *et al.* [1].

Not only did the ambient heuristics do better than Nielsen's in a domain for which Nielsen's heuristics *were not* intended, they performed equally well to reported performance of Nielsen's heuristics in domains for which they *were* intended. In particular, the canonical measure of a successful heuristic evaluation is that 3-5 evaluators can find 40-60% of issues with a display, a measure met by our ambient heuristics.

A next step for this work is to apply our final set of heuristics to a number of ambient displays, including the daylight display and busMobile. Our plan is to revise those displays based on the heuristic evaluation results, and compare the usability of the revised displays to the original displays using an expensive summative technique such as a longitudinal study.

This paper presents a working set of ambient display heuristics. As we continue to use them, we expect our heuristics to evolve, just as Nielsen's heuristics have evolved over time. Additionally, these heuristics represent an initial step toward establishing ambient display design guidelines. Lastly, we hope to develop a wider range of formative and summative techniques for evaluating ambient display usability.

Acknowledgments

Many students contributed to this work, including Lisa Chan, Chinmayi Bettadapur, Steven Chan, and Scott Carter. Thanks as well to the participants in our study for their time and effort.

REFERENCES

1. K. Baker, *et al.* Empirical development of a heuristic evaluation methodology for shared workspace groupware. In *Proc. of CSCW 2002*, pp. 96-105. ACM Press, 2002.
2. C. Chafe and G. Niemeyer. Oxygen flute. Installation by Artists, San Jose Museum of Art, October 13-2001 - June, 2002.
3. A. Chang, *et al.* Lumitouch: An emotional communication device. In *Proc. of CHI 2001*, pp. 313-314. ACM Press, 2001.
4. C. Chewar and D. S. McCrickard. Adapting UEMS for notification systems. Accepted to Ubicomp 2002 Workshop 9, September 2002.
5. W. D. Gray and M. C. Salzman. Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction*, 13(3):203-261, 1998.
6. L. Hallnäs and J. Redström. Slow technology; designing for reflection. *Personal and Ubiquitous Computing*, 5(3):201-212, 2001.
7. J. M. Heiner *et al.* The information percolator: Ambient information display in a decorative object. In *Proc. of UIST 1999*, pp. 141-148. ACM Press, 1999.
8. H. Ishii, *et al.* Pinwheels: Visualizing information flow in architectural space. In *Proc. of CHI 2001*, pp. 111-112. ACM Press, 2001.
9. H. Ishii and B. Ullmer. Tangible bits: Towards seamless interfaces between people, bits and atoms. In *Proc. of CHI 1997*. ACM Press, 1997.
10. H. Ishii, *et al.* Ambient displays: Turning architectural space into an interface between people and digital information. In *Proc. of CoBuild 1998*,. Springer-Verlag, 1998.
11. E. H. M. van Dantzich, *et al.* Scope: Providing awareness of multiple notifications at a glance. In *Proc. of AVI 2002*. ACM Press, 2002.
12. B. MacIntyre, *et al.* Support for multitasking and background awareness using interactive peripheral displays. In *Proc. of UIST 2001*, pp. 41-50. ACM Press, 2001.
13. E. Mynatt, *et al.* Designing Audio Aura. In *Proc. of CHI 1998*. ACM Press, 1998.
14. E. Mynatt, *et al.* Digital family portraits: Providing peace of mind for extended family members. In *Proc. of CHI 2001*, pp. 333-340. ACM Press, 2001.
15. J. Nielsen. Ten usability heuristics. http://www.useit.com/papers/heuristic/heuristic_list.html, accessed on August 20, 2002.
16. J. Nielsen. Noncommand user interfaces. *Communications of the ACM*, 36(4):83-99, Apr. 1993.
17. J. Nielsen and T. K. Landauer. A mathematical model of the finding of usability problems. In *Proc. of ACM INTERCHI 1993*, pp. 206-213, 1993.
18. J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. In *Proc of CHI 1990*, pp. 249-256. ACM Press, 1990.
19. J. Nielsen and the Nielsen Norman Group. Internet: <http://www.useit.com/papers/heuristic/>, 2002.
20. D. Tesselndorf, *et al.* An ordering of secondary task display attributes. In *Proc. of CHI 2002*, pp. 600-601. ACM Press, 2002.
21. K. Vredenburg, *et al.* A survey of user-centered design practice. In *Proc. of CHI 2002*, pp. 471-478. ACM Press, 2002.
22. M. Weiser and J. S. Brown. Designing calm technology. Internet: <http://www.ubiq.com/weiser/calmtech/calmtech.htm>, 1995.
23. T. White and D. Small. An interactive poetic garden. In *Proc. of CHI 1998*. ACM Press, 1998.
24. A. Woolrych and G. Cockton. Why and when five test users aren't enough. In *Proceedings of IHM-HCI 2001 Conference*, volume 2, pp. 105-108, 2001.