

Context of Use Evaluation of Peripheral Displays (CUEPD)

N. Sadat Shami, Gilly Leshed, and David Klein

Cornell University Information Science Program,
301 College Ave., Ithaca, NY 14850, USA
{sadat, gilly, dik4}@cornell.edu

Abstract. A gap exists between the growing prevalence of peripheral displays and appropriate methods for their evaluation. Mankoff et al. [11] present one attempt to bridge this gap by adapting Nielsen's Heuristic evaluation to the defining characteristics and goals of peripheral displays. In this paper, we present a complementary approach that depends on active user participation and emphasizes the experience of using peripheral displays. The Context of Use Evaluation of Peripheral Displays (CUEPD) captures context of use through individualized scenario building, enactment and reflection. We illustrate the CUEPD method in a study to evaluate two peripheral displays. The evaluation using CUEPD revealed important design recommendations, suggesting that the method may be an important advance in evaluation methods for peripheral displays.

1 Introduction

Increasingly, information is being displayed in the periphery of our attention. These *peripheral displays* are designed to move back and forth between the center and periphery of a user's attention since they require minimal cognitive processing. As a result, they allow users to focus on a primary task while maintaining opportunistic awareness of a secondary task [15]. McCrickard et al. suggest a classification of these types of displays, according to three critical parameters – interruption, reaction and comprehension (IRC) [13]. Each display receives a rating between 0 and 1 on each of the three IRC parameters. For example, an onboard vehicle navigation system invokes moment to moment reaction by prompting turns along the road without causing interruption or requiring deep comprehension about the route. As such, it will receive an IRC rating of (010). An instant messaging program notifies users of arrival of new email and when a contact logs in. A user can then attend to important emails or communicate with a contact by redirecting activity as necessary. Such an instant messaging program would receive an IRC rating of (110). In this paper, we focus on the set of displays that receive ratings of 0 on interruption and between 0 and 1 on both reaction and comprehension.

The research on peripheral displays has focused primarily on their design rather than on methods of evaluating them [11]. These displays are difficult to evaluate [1,11,13]. Traditional HCI evaluation methods focus on productivity oriented metrics such as successful task completion rates, time taken for completion, error rates etc. It

is difficult to use these metrics for measuring perception and attention [10,11]. Furthermore, traditional evaluation methods focus on how successfully a system supports a user's primary task. But peripheral displays are about secondary tasks. For example, results obtained from a study that had users simply use a vehicle navigation system as a primary task would not be as meaningful as a study that had users drive while using the navigation system. This difference in the nature of peripheral displays compared to other systems highlights the need for an evaluation method tailored for peripheral displays.

Most of the designs of peripheral displays published in the literature have been evaluated through ethnographic methods [2,8,9], some through controlled experiments [12,18], others by a combination of the above [3], while some have not been evaluated at all [4,5,7,17]. To the best of our knowledge, only one method has been explicitly developed to evaluate peripheral displays. Mankoff et al. extend Nielsen's Heuristic Evaluation by modifying his set of heuristics to fit peripheral displays. Their discount evaluation method provides guidance during the early stages of design by reminding designers about usability principles [11]. Typically, different evaluation methods are used at different stages of the design process [14]. In this paper, we introduce the Context of Use Evaluation of Peripheral Displays (CUEPD – *pronounced cupid*), which can be used when designers have at least a working prototype and are interested in improving its future design. As such, CUEPD is complementary to Mankoff et al.'s method as it is used later on in the design process.

2 Context of Use Evaluation of Peripheral Displays (CUEPD)

As one of two research challenges facing the ubiquitous computing field, Abowd & Mynatt emphasize the importance of developing evaluation methods that capture the context in which a system is used [1]. Context of use refers to the setting in which a display is employed and it is assumed to influence our attention and perception. For example, the human 'spotlight of attention' may or may not focus on the information a display attempts to convey because of its surrounding context. It is noteworthy that moving from a primary to a secondary task and back is not only about switching visual focus, but also about the cognitive processing necessary to perform such action.

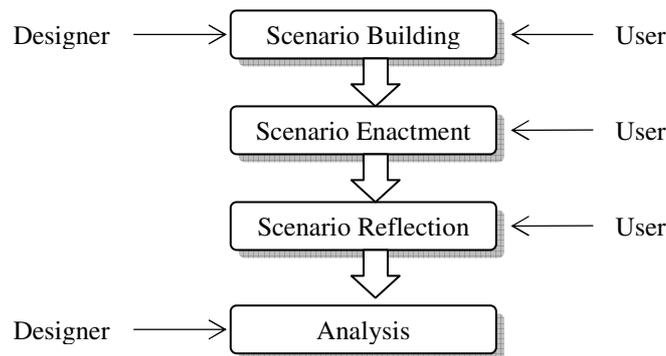


Fig. 1. Conceptual model of CUEPD

We describe the stages of CUEPD below, explaining how context of use is captured at each stage. A conceptual model of CUEPD is displayed in Figure 1. We stress the fact that users need not be familiar with the evaluation method or have evaluation experience to participate. The only requirement is that they be potential users of the display.

2.1 Scenario Building

The first step of CUEPD is a question and answer session between the designer and a user with the purpose of creating a scenario that captures how the information presented in the display is accessed. The designer asks questions on how the user accesses the *information* presented in the display, rather than the display itself. For example, in an evaluation of a peripheral network monitor displayed on a computer desktop, the designer would ask a system administrator about ways she monitors network traffic. Based on the answers that the user provides, the designer creates a scenario of use that the user will act out. Within the scenario, the designer will identify one or more primary tasks for the user to complete while providing the opportunity to access the peripheral display. In the network monitor evaluation example, the designer could create a scenario where installing software patches and creating new accounts on the system are primary tasks, and keeping track of network traffic is the secondary task. This manner of scenario building is a departure from traditional scenario-based design since a designer and user collaboratively construct a scenario, as opposed to designers doing it themselves. Working together with the user to build a scenario allows the peripheral display to be evaluated in the context of how an individual would use it in a real situation.

2.2 Scenario Enactment

Once the scenario has been developed and described to the user, the user performs the primary task while having the opportunity to access the display. Note that accessing the display is never the primary task. In fact, during the primary task, the peripheral display should not be at the center of a user's attention. However, once the peripheral display is noticed and accessed, it then moves from the periphery of a user's attention to the center. Evaluating if and how users use a peripheral display within the context of a primary task allows designers to determine how easily the display moves back and forth between the periphery and center of a user's attention.

2.3 Scenario Reflection

After acting out the scenario, users are given a 10 item questionnaire to fill out. The questionnaire design was informed by our literature review of defining attributes of peripheral displays. We believe the best way to get users to articulate strengths and weaknesses of the design of a peripheral display are to frame questions based on its defining attributes. Mankoff et al. took a similar approach when they focused on definitional attributes of peripheral displays to modify Nielsen's Heuristics [11]. We used an iterative approach to determine the questions on the questionnaire. Certain

questions were modified based on initial responses from participants. We identified five categories representing definitional attributes of peripheral displays, and developed questions that address our categories, as listed in Table 1. The categories reflect the trade-offs in designing peripheral displays. For example, a peripheral display should be *noticeable* and allow for *division of attention* at the same time. It should provide *comprehensible* and *relevant* information yet be *engaging*. Rating high on all these categories is a challenge for designers.

Table 1. Questionnaire questions grouped by category

Noticeability Q 1. Did you notice the display? Q 7. While performing your primary task were you aware of the opportunity to access the display?
Comprehension Q 2. Were you able to understand the information in the display? Q 5. Were you able to understand the information just by glancing at it?
Relevance Q 3. Did the display provide you the information you needed?
Division of attention Q 4. Was the display located outside the focus of your attention? Q 6. Were you able to adequately focus on your primary task? Q 8. Were you able to shift your attention between your primary task and the display smoothly?
Engagement Q 9. Did you find the design of the display attractive? Q 10. Did you enjoy using the display?

Each question in our questionnaire was followed by a ‘Yes/No’ option and space for an open-ended response. By asking users to choose between ‘Yes’ and ‘No’ we attempted to give designers clear feedback regarding aspects of design, while at the same time allowing participants to express themselves in the open-ended portion of the response. In the end, we found that the explanations were the most useful part of the questionnaire because they provided more information about how the display mapped onto the user experience.

2.4 Analysis

The analysis focuses on how well the display addresses the five categories described above. For each question, the designer assigns 1 for a positive and 0 for a negative response, based on the ‘Yes/No’ answer and explanation provided in the open-ended response. Next the numeric values are averaged in the following way for each category:

$$\frac{\text{positive responses}}{\text{questions in category} \times \text{users}} \quad (1)$$

A high or low rating is an early indicator of which category's open ended responses should be looked at first. Those explanations then allow for in-depth interpretation regarding the user's experience. The numeric values also provide approximations on how well the display performs on the five categories.

3 Testing the CUEPD Method

We subsequently tested CUEPD on real peripheral displays. Gray and Salzman argue that comparative studies of evaluation methods often suffer validity weaknesses, following different settings and metrics measured by each method [6]. As a result, relative effectiveness of one method over the other can hardly be trusted. Therefore, we tested our method independently to see whether it was effective in capturing context of use in peripheral displays.

3.1 Displays Used

We evaluated two displays, which are both relatively common. One was a peripheral display for a computer desktop called 'Weather Watcher' (a free download at <http://www.singerscreations.com/>), as shown in Figure 2. The other display was a stock ticker located along the entrance corridor of the Business School of a large northeastern university (Figure 3). We chose these two because we wanted to check the validity of CUEPD on both 'virtual' and 'tangible' peripheral displays. Our focus was not to evaluate these displays *per se*, but rather to determine whether CUEPD was effective in evaluating them. The Weather Watcher and stock ticker both act as passive displays providing information that needs to be comprehended without interrupting users and sometimes requiring them to react to the information. Consequently, they would both receive IRC rankings of (0/0.5/1) in the context of our evaluation.



Fig. 2. Weather Watcher display showing 31 degrees



Fig. 3. The stock ticker is above the TV screens

3.2 Participants

The participants we recruited were undergraduate, graduate and MBA students of the same large northeastern university. They received a \$10 gift certificate in exchange for their participation. 95% of participants had been using computers and the internet for more than 5 years. In the Weather Watcher evaluation, we had 7 undergraduate and 3 graduate students, with 80% having little or no self reported experience evaluating a system/software, and 20% having between 1 and 3 years of experience. In the stock ticker evaluation, we had 9 MBA students and 1 fourth year undergraduate business student, with 60% having no evaluation experience at all and 40% having between 1 and 3 years of experience. None of the Weather Watcher participants were previously familiar with the software while the business students were acquainted with the stock ticker at the entrance to their building.

3.3 Procedure and Task

The Weather Watcher evaluation was conducted on a laptop computer that had the software installed on it. The stock ticker evaluation took place near the location of the stock ticker at the Business School. After filling out a pre-task questionnaire on demographics and prior evaluation experience, participants were guided through the stages of CUEPD. Typical scenarios for the Weather Watcher included checking email and reading online news as primary tasks. Typical scenarios for the stock ticker included entering the business school to go to class or to the library as primary tasks. Accessing information from the displays was a secondary task in all these scenarios. With permission of participants, the scenario building conversation was recorded. Participants then acted out the scenario and completed our questionnaire.

4 Results

Table 2 presents the category ratings for the Weather Watcher and stock ticker using CUEPD. To verify the reliability of the results, the questionnaire data was coded independently by two coders. Inter-coder reliability using Cohen's Kappa was a satisfactory 0.905 and 0.865 for Weather Watcher and stock ticker evaluations respectively.

Table 2. Weather Watcher and Stock Ticker ratings

Category	Weather Watcher	Stock Ticker
Noticeability	71%	62%
Comprehension	67%	69%
Relevance	56%	13%
Division of attention	80%	63%
Engagement	60%	73%

We report how CUEPD allowed us to arrive at design recommendations. From Table 2, we see Weather Watcher received the highest favorable rating on *division of attention*. That alerted us to look at the open ended responses in that category to determine why. One Weather Watcher participant reported that it was “small enough not to distract a lot but big enough to be readable and noticeable.” Since *division of attention* involves a trade-off with *noticeability*, we then read explanations in that category. We found participants reporting that “I forgot it was there” and “it did not ‘jump’ at me”, which might explain why it rated high in *division of attention* but lower in *noticeability*.

For the stock ticker, we see that it received its lowest rating on *relevance*. According to the open ended responses in that category, users reported that “the individual company stickers were pretty random” and lacked “overall levels of indexes”. It also ranked relatively low in *noticeability*. Participants reported that it was located “too high.” As location is part of context, CUEPD was successful in creating a contextual setting that facilitated such responses.

To improve *noticeability* in the Weather Watcher, we suggest adding a meteorological symbol in the background of the temperature figure, i.e. a cloud or the sun, to emphasize the relation of the number to weather information. The *relevance* of the stock ticker can be improved by presenting both specific stock information in a predetermined order and general trends/indexes of the market. Suggestions for *noticeability* include relocating the display lower and in front of the entrance rather than on the side wall. These suggestions illustrate how CUEPD provides practical design recommendations for peripheral displays.

5 Discussion

Our study suggests that CUEPD is a functional evaluation method for peripheral displays that captures context of real use through scenario building, enactment, and reflection. Rosson & Carroll describe the use of scenarios as a basis for usability evaluation [16]. We take this approach further, allowing for potential individual differences and preferences by customizing the scenarios to each user. This ‘participatory’ evaluation advances traditional scenario-based methods as the designer and user collaboratively construct the scenario. If recorded, as we did in our study, a repository of scenarios is produced, facilitating the understanding of typical uses, as well as unintended ones. Under CUEPD, there is a possibility that each participant might receive a different scenario, which we believe would produce a set of valuable information for the designer. In particular, the different scenarios would qualify the numeric ratings, and illustrate the importance of the qualitative responses in evaluating a display within its context of use. In that sense, CUEPD is in the direction of evaluation methods that go beyond usability and attempt to measure experience.

The primary vs. secondary task is an important distinction for capturing context of use. For example, Plaue et al. successfully evaluated recall of data presented in peripheral displays in a *primary task* setting [15]. In contrast, CUEPD requires users to focus on a task apart from the display, which enables an evaluation of the ease of switching between primary and secondary tasks, an important characteristic of peripheral displays. Maglio & Campbell evaluated scrolling text displays through a

series of experiments using a primary task/secondary task distinction [12]. However, their study was not intended to be a formal evaluation method for peripheral displays.

Since time to market is a critical factor nowadays, CUEPD is useful in eliciting valuable user feedback without requiring a lot of time commitment. It took on average half an hour for us to conduct our study with each participant. Furthermore, CUEPD can be easily applied with average users that have little or no evaluation experience. The majority of participants in our study did not have substantial evaluation experience. With increasing demand for discount evaluation methods, CUEPD is similar to the endeavor of Mankoff et al. to produce such methods [11].

6 Conclusion and Future Work

In this study, we combined an approach of tailoring an existing evaluation method for specific systems, e.g. in [11], with the call to develop evaluation methods that capture context of use [1]. The result, CUEPD, can either be used to evaluate already deployed systems, such as the stock ticker, or working prototypes, as demonstrated by the Weather Watcher. We fully expect that as our method is used and tested, it will evolve, as other evaluation methods have evolved. While we have focused on McCrickard's [13] parameters of reaction and comprehension, future work could extend to displays having high levels of interruption as well.

Designers select methods for evaluating their design from a repository of evaluation methods [14]. Our method serves as a useful addition to the evaluation methods designers of peripheral displays have at their disposal.

Acknowledgements

This research was supported by the funding of Jeff Hancock. We thank Kirsten Boehner, Nathan Bos, Matt Brochstein, Anind Dey, Geri Gay, Jeff Hancock, Elaine Huang, and Joseph Kaye for their valuable comments on previous drafts.

References

1. Abowd, G.D., Mynatt, E.D. Charting past, present, and future research in ubiquitous computing. *ACM Trans. Comp.-Hum Inter.* 7, 1, (2000), 29-58.
2. Cadiz, J. J., Venolia, G., Jancke, G., Gupta, A. Designing and deploying an information awareness interface. In *Proc. CSCW 2002*, ACM Press (2002).
3. Denoue, L., Nelson, L., Churchill, E. AttrActive windows: dynamic windows for digital bulletin boards. *Ext. Abstracts CHI 2003*, ACM Press (2003).
4. Dieberger, A. Supporting collaboration through passing informal notes to peripheral displays. *Ext. Abstracts CHI 2002*, ACM Press (2002).
5. Gaver, B. Provocative Awareness. *Computer Supported Cooperative Work*, 11, (2002), 475-493.
6. Gray, W.D., Salzman, M.C. Damaged Merchandise? A review of experiments that compare usability evaluation methods. *Hum. Comp. Inter.* 13, (1998), 203-261

7. Heiner, J., Hudson, S., and Tanaka, K. The Information Percolator: Ambient Information Display in a Decorative Object. In *Proc. UIST 1999*, ACM Press (1999), 141-148.
8. Huang, E. M., Mynatt, E. D. Semi-public displays for small, co-located groups. In *Proc. CHI 2003*, ACM Press (2003).
9. Huang, E. M., Tullio, J., Costa, T. J., McCarthy, J. F. Promoting awareness of work activities through peripheral displays. *Ext. Abstracts CHI 2002*, ACM Press (2002).
10. Karat, J. Beyond Task Completion: Evaluation of Affective Components of Use. In J. Jacko and A. Sears (Eds.), *The Human Computer Interaction Handbook*. LEA, 2003.
11. Mankoff, J., Dey, A. K., Hsieh, G., Kientz, J., Lederer, S., Ames, M. Heuristic evaluation of ambient displays. In *Proc. CHI 2003*, ACM Press (2003).
12. Maglio, P. P., Campbell, C. S. Tradeoffs in displaying peripheral information. In *Proc. CHI 2000*, ACM Press (2000).
13. McCrickard, D. S., Chewar, C.M., Somervell, J.P., Ndiwalana, A. A model for notification systems evaluation – Assessing user goals for multitasking activity. *ACM Trans. Comp.-Hum. Inter.* 10, 4, (2003).
14. Olson, J. S., Moran, T. P. Mapping the method muddle: Guidance in using methods for user interface design. In M. Rudisill, C. Lewis, P. B., Polson, and T. D. McKay, (Eds.), *Human-Computer Interface Design: Success Stories, Emerging Methods and Real-World Context*, Morgan Kaufmann, San Mateo, Calif, 1996.
15. Plaue, C., Miller, T., Stasko, J. Is a picture worth a thousand words? An evaluation of information awareness displays. In *Proc. Graphics interface (2004)*.
16. Rosson, M., and Carroll, J., M. Scenario-based design. In J. Jacko and A. Sears (Eds.), *The Hum. Comp. Inter. Handbook*. LEA, 2003.
17. Skog, T. Activity wallpaper: ambient visualization of activity information. In *Proc. DIS 2004*, ACM Press (2004).
18. Tyman, J., Huang, E. M. Intuitive visualizations for presence and recency information for ambient displays. *Ext. Abstracts CHI 2003*, ACM Press (2003).