Measuring Distraction and Awareness
Caused by Graphical and Textual Displays in the Periphery

Jacob Somervell, Ragavan Srinivasan, Kim Woods, Omar Vasnaik
Computer Science Department
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061
{jsomerve,rsriniva,kwoods,ovasnaik}@vt.edu

Abstract - Peripheral displays provide a means to present information to users in the periphery of the computer desktop. There are several forms of peripheral displays, some of the most widely recognized being stock tickers, email alert tools, and percent-done indicators. The use of such displays inevitably leads to the questions of how successful they are at communicating information to the user and how distracting they are when performing other tasks. This paper presents a study on the effectiveness of using peripheral displays of varying type and number to convey potentially interesting but non-urgent information to the user. The participants in this study were asked to perform a primary task while one or more peripheral displays disseminated information. The results of the study seem to indicate that both textual and graphical peripheral displays impact performance on primary tasks yet both can effectively communicate information.

Keywords: human-computer interaction, awareness, peripheral displays, peripheral tasks, evaluation, graphical display, textual display, ticker

1. Introduction

People are curious by nature. As a result of this curiosity, we, as humans, have developed a constant need for information. One result from this need for information was the evolution of informational displays. Ranging from billboards to televisions, we use informational displays to bring us the information that we seek. Peripheral displays are a subset of informational displays that reside in the periphery of the user. They only come into the user's primary attention when certain events occur that catch the user's attention or when the user explicitly looks at the display.

It should be understood that peripheral displays are not, and should not be limited to those that reside on a computer screen -- peripheral displays can take on many forms. Some examples of peripheral displays include stock tickers, road signs, billboards, clocks, and windows.

More recently, new techniques for displaying information in the periphery have emerged, including audio cues [4,7], bubbling machines [5], and even dangling strings hanging from ceilings [8]! Web-based companies have gotten into the act as well – sites like Yahoo (yahoo.com) and ESPN (espn.com) offer peripheral displays for the computer desktop. These peripheral displays exist to augment the users' awareness of information ranging from keeping track of stock prices to alerting us of new email to determining if network traffic is heavy, to pretty much almost anything a user desires.

Many people have multiple peripheral displays running on their machines at any given time. People use systems clocks, load monitors, e-mail monitors, stock tickers, and a variety of others. How does having these peripheral displays affect the person's ability to get work done? Does presenting the information in a graphical way become more effective than using text? Which types of displays would be more effective, graphical or textual? Are graphical displays more distracting than textual ones? Our goal is to attempt to understand the answers to these questions experimentally.

To address these questions, we conducted an experiment in which participants performed a browsing task while simultaneously monitoring a peripheral display (or displays). We examined whether graphical displays like gauges and percent-done indicators might be better than text-based displays like stock tickers and faders in terms of the amount of distraction and interference experienced by the user. In addition, we explored the effects using multiple peripheral displays may have on our ability to accomplish simple tasks. In so doing, our intent was to test how increasing the number of peripheral displays would affect user performance.

The rest of this paper contains some information on related work in the field, a description of the experiment that was performed, the results of the experiment, some discussion on what the results imply, and finally some directions for future research in the area of peripheral displays.

2. Related work

While there has been recent work in the development of peripheral displays, little work has been done in the

evaluation of them. The most relevant results are discussed here.

McCrickard et al. studied how various textual animations in the periphery help maintain information awareness while performing browsing tasks [6]. They found that peripheral displays did not have a significant impact on user task performance on a browsing task. Maglio et al., on the other hand, ran similar experiments but used more cognitive intensive tasks; and they found that having peripheral displays did have a significant impact on user task performance [5]. In both of these studies the type of peripheral display was textual in nature with the information updated through some form of animation

McCrickard et al. used three types of animation: fade, blast, and ticker [6]. The fade display changed the information by changing the text font from the background color to the desired foreground color (typically white to black). The blast display instantaneously changed the information in the display. The ticker display works in the same way most people are accustomed to: information is scrolled from right to left across the display, in this case updating one pixel at a time. User performance was measured in browse time – the amount of time to complete simple browsing tasks. They found that the presence of peripheral displays had no effect on user performance, irregardless of whether the display was a fade, blast, or ticker.

Maglio et al. also used ticker displays in their study. Specifically, they were studying whether continuous scrolling, discrete scrolling, or serial presentation were better at communicating information, and how well the information was remembered while performing a document editing task [5]. Continuous scrolling, as the name implies, is text scrolling from the right to the left. Discrete scrolling is a little different. The information scrolls in from the side, stops moving, then scrolls off the side. Serial presentation works somewhat like the blast display. The information simply appears in the display with no animation. Maglio et al. measured user performance with the number of corrections made to the document and with responses to recognition tests. They found that the presence of the displays did impact user performance.

The different results from Maglio and McCrickard that were mentioned do not seem to come from the type of display. Instead, it seems the main difference in these studies was the type of primary task. McCrickard et al. used a browsing task, similar to the one described below, whereas Maglio et al. used a document-editing task. This difference in task type could account for the seemingly contradictory results between the two studies.

In a related study, Czerwinski et al. explored the effects of instant messaging on user task performance [1,2]. The participants were asked to search for specific titles in a large database of listings. While doing this search task, one of the experimenters would send an instant message using Microsoft Instant Messenger. Search time and reaction time

(to the instant message) were used to measure performance. Their work indicates that these types of interruptions (instant messages) have negative effects on user performance. The type of task Czerwinski et al. used was a type of search task; here again a more cognitively intense type of task, which again most likely explains why these results were different from those of McCrickard et al.

3. Experimental setup

The purpose of this experiment was to examine the relative advantages of textual and graphical peripheral displays and to determine the effects of multiple peripheral displays on task performance. The primary task was a browsing task—browse through some information looking for answers to specific questions. This browsing task was a modified version of the one used in [6]. Each experiment consisted of six rounds of questions, four questions per round. The goal was to answer the questions as quickly as possible. All answers were in numerical format to reduce typing. The information was displayed in a simple browser, similar to existing web browsers like Netscape and Internet Explorer. 'Back' and 'Forward' buttons were provided to aid navigation through various links dispersed within the information (figure 1). Links were designated in the traditional way of using blue, underlined text. Once the participant found the answer, they typed it into the answer box and clicked the "OK" button. This ended that browsing task and prompted the next browsing task to appear. Incorrect answers produced a "beep" and the participant was then required to try again until the correct answer was supplied.

Four test groups were designed to measure user performance on the browsing task while the user was also trying to complete a given number of awareness tasks. The awareness tasks relied on information that was displayed in a textual display, a graphical display, or both. These awareness tasks are representative of some of the peripheral tasks that people actually do while browsing. People often monitor stock quotes, news headlines, weather, and sports scores while they work. It is also common for people to be using their load monitors, performance monitors, or percent-done indicators while running important processes. The tasks used in this experiment are similar in nature to these. The setups for each experimental test group are described below.

3.1. Control group

The control group simply had to complete the browsing tasks; there were no peripheral displays shown while they worked on answering the questions. Thus we could determine a base time for completing the browsing tasks. Figure 1 illustrates the setup for this test group. The previous description of the browsing task covers the setup for the control group.
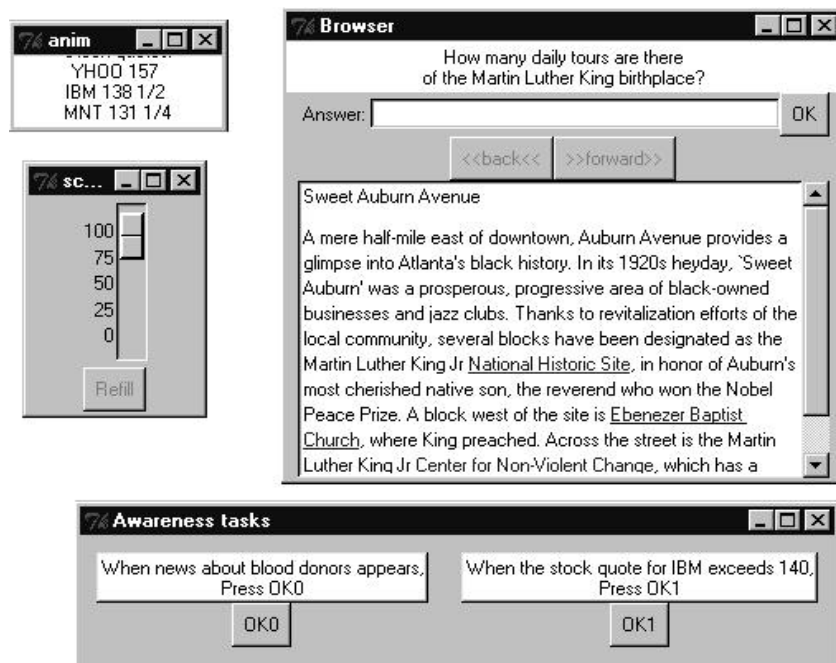
Figure 1. This is the interface used in these experiments. The main browsing task was done in the larger window on the right. The control group only saw this part of the interface. The browsing task appears at the top of the window. The answers are in the text in the browse window. The scale group saw the browse window and the scale to the immediate left. Clicking refill completed the scale awareness task. The fade/ticker group saw the browse window and the awareness tasks at the bottom, along with the display at the upper left. The awareness tasks could be completed when the relevant information was shown in the textual display. The fade/ticker scale combo group saw all of the displays together.

3.2. Fade/Ticker group

The first experimental test group included two simple awareness tasks per round in addition to the browsing tasks. Basically for each round there were two awareness tasks the user was asked to complete in addition to the previously described browsing tasks. These awareness tasks depended on information that was displayed in a peripheral display. Fade and ticker displays were used to present the auxiliary information from which the awareness tasks could be completed. These displays were taken directly from McCrickard et al. [6]. Both displays presented information on stock prices, weather, sports, and news stories. When the information appeared in the display, the user was asked to press a button to indicate he/she saw the information. An example question would be: "When the stock price of IBM reaches 140 press OK1". In the peripheral display, information containing stock quotes, among other types of information, would appear. A single button press was chosen because it mimics what might happen when a person becomes aware of some certain information. An example is if the temperature drops below 40, a person may turn on a heater. Figure 1 illustrates the setup used in this test group.

3.3. Scale group

The second experimental group was similar to the previously described setup but with one fundamental change. The peripheral display and the corresponding awareness tasks were removed and replaced with a new awareness task. This new task involved a different type of peripheral display. The display was a numerical scale that ranged from 100 down to zero. Starting from 100, the scale would slowly decrease in value, stopping at zero. A "refill" button was provided, which when clicked would reset the scale value to 100. The task associated with this display was to click the "refill" button after the scale fell below 25. This task was to be completed once per round. The rate at which the scale decreased depended on the round the participant was in. For example, if the user was working in round one, the scale took 90 seconds to go down to the 25 value; round two took 100 seconds, three took 85, four took 105, five took 95, and six took 110. Different values were used so that the participant could not predict the amount of time it would take for the scale to reach the desired level (below 25). The specific choices for the times are based on data from the McCrickard experiment [6]. There, participants took at least 120 seconds to complete a round. The intent was to have the scale actually reach the desired

level before the participant completed the round of questions. See figure 1 for an example of the setup.

This particular choice of display mimics numerous peripheral tasks a user might do. The scale display is similar to a fuel gauge in automobiles. One drives for some amount of time and then must refill their tank with fuel. The "refill" button associated with the scale display simulates this refilling of the tank. Similarly, while downloading a file one might want to have a percent-done indicator on their screen that informs them when a large file will complete its download. In addition, it is not uncommon for people to have load monitors running while they work. If the monitor spikes, the user might kill a process. The scale design emulates these activities.

3.4. Fade/Ticker scale combo group
The third experimental test group was asked to complete the browsing tasks and both types of awareness tasks described earlier in the Fade/Ticker and Scale groups. This means the participants had to keep track of the peripheral display that had textual information (fade or ticker) and they had to keep track of the scale, in addition to performing the browsing tasks. We placed the scale in the same place that it occurred in the Scale test group so that the placement of the displays in the interface would not cause statistically significant results (figure 1). The section on future work at the end of this paper discusses different possibilities in placement setups.

4. Participants
Twenty-nine people were included in this experiment. Both students and non-students were used. All participants were volunteers, and thus were not compensated for their time. Participants had varying degrees of computer skills but each was familiar with the web browser concept: hypertext links with 'back' and 'forward' buttons to facilitate navigation. The experiments were conducted on IBM compatible machines with either Windows 98 or NT operating systems. The interface was created in Tcl/Tk and the experiment was run using Wish 8.3. Tests were performed in isolation with one participant per computer. The experiment was explained to each participant both verbally and electronically, with examples on the computer to illustrate the idea.

5. Results
This section and the following one describe the results and observations we made from our experiments. Data were collected as the participants interacted with the interface. Various measurements were taken to discern user performance on the browsing task. The following is a description of these measurements.

Browse time: This is the time a participant took to finish all primary browsing related tasks for each round. This was taken as the average time a participant spent on a round, per test group.

Reaction time: This is the time that elapsed between the instant a participant could react to an event (e.g. the scale falling below the 25 mark) and the instant he actually reacted to the event (e.g. clicking the 'refill' button). This was taken as the average reaction time, per test group.

Task completion rate: This is the percentage of awareness tasks that participants completed in the experiment per test group. <u>*Note*</u>: *All the measurements of time were in seconds and were measured using the system clock for accuracy.*

After collecting the aforementioned data, we carried out ANOVA (Analysis of Variance) tests on it to determine any statistical significance. ANOVA is a valid, standard and accepted statistical technique for testing differences among group means. If an ANOVA indicates a significant difference, the t-test can then be used to test differences between pairs of means.

5.1 Measurements on browse times:
The mean and variance values of browse times for each of the test groups are tabulated in table 1.

Group	Mean Browse time	Variance
Control	163.1296	3971.429
Fade/Ticker	287.3333	1758.367
Scale	250.5000	6891.211
Fade/Ticker Scale Combo	259.9375	9437.904

Table 1: Mean and variance values of browse times for each participant group.
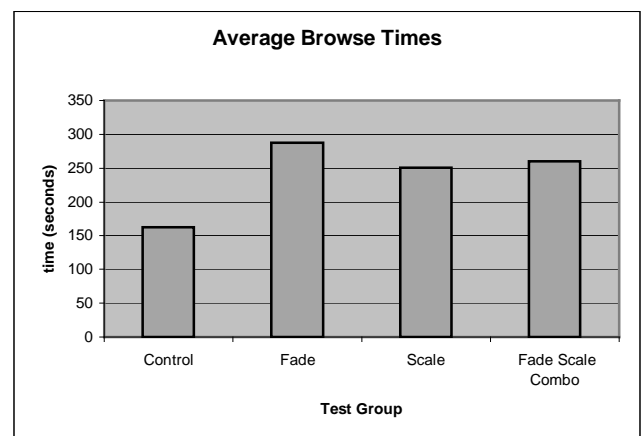
Figure 2: The chart is a representation of the average browse times for the four participant groups. It shows a significant difference between the average browse time of the Control group and all the other groups. It also shows there is not a significant difference in the average browse times between the Fade/Ticker, Scale, and the Fade/Ticker Scale Combo groups.

The ANOVA test on the above values resulted in a p-value of 0.0171 between groups, which indicated a significant difference. (See Figure 2 for graphical representation of browse times.) This necessitated further pairwise t-tests on the values, the results of which are presented in table 2.

Pairs of Groups	p-values
Control, Fade/Ticker	**0.0005**
Control, Scale	**0.0185**
Control, Fade/Ticker Scale Combo	**0.0131**
Fade/Ticker, Scale	0.1773
Fade/Ticker, Fade/Ticker Scale Combo	0.2664
Scale, Fade/Ticker Scale Combo	0.4259

Table 2: Results of pairwise t-tests for browsing time. Bold values indicate statistical significance

As seen in table 2, there was a significant difference in the browse times between each of the test groups and the control group. The implications of this are covered in the discussion section below.

5.2 Measurements on reaction times:
The mean values of reaction times for each of the participant groups that had a graphical (scale) display are presented in the Table 3 and Figure 3 below:

Group	Mean value (in seconds)
Scale	14.05556
Fade/Ticker Scale Combo	16.02667

Table 3: Mean values of reaction times for the scale and combo groups.
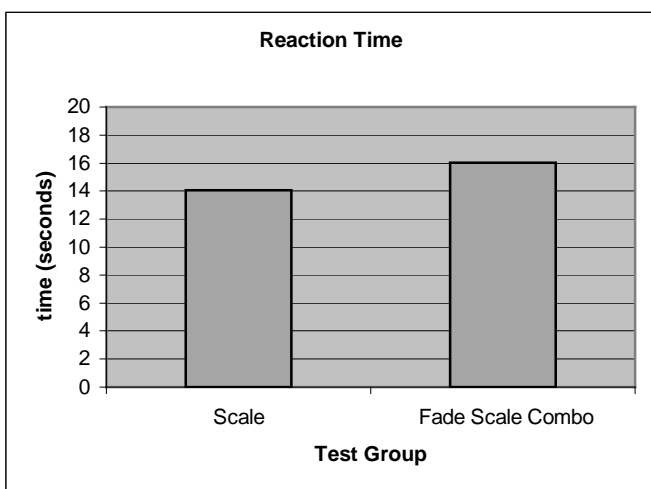
Figure 3: Graphical representation of the mean reaction times for the Scale and Fade/Ticker Scale Combo groups.

The ANOVA test on these values resulted in a p-value of 0.2630 between groups, which indicated no significant difference. Thus, further pairwise t-tests were not necessary.

5.3 Measurements on task completion rates:
The following Table and figure 4 contain the mean and variance values for task completion rates:

Group	Mean value (in %)	Variance
Fade/Ticker	91.6667	0.0167
Scale	91.6667	0.0194
Fade/Ticker Scale Combo	78.4722	0.0286

Table 4: Task completion rates for the various groups.

The ANOVA test on these values yielded a p-value of 0.1852 between groups, which indicated there was no significant difference, and also eliminated the need for further pairwise t-tests.

The following section will discuss the meanings of the various results that have been observed and tabulated in this section.

<u>Note</u>: As in the case of the reaction time measurements, in the case of the control group, there were no extra tasks, and hence, no measurements. Further, it is interesting to note the number of tasks for each group. Remember, each group had 6 rounds of experiments.
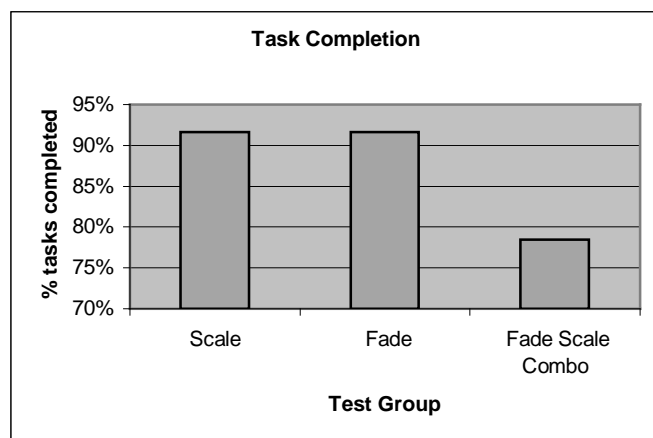
Figure 4: This chart is the representation of the average task completion rates of the Scale, Fade/Ticker, and Fade/Ticker Scale Combo groups. Though the ANOVA tests do not yield a statistically significant value, they do suggest a strong trend that increasing the number of displays does affect the task completion rates, as suggested in the above chart.

6. Discussion

This section presents a discussion of the results. There are several interesting items that stem from this experiment.

6.1 It takes significantly longer to complete a primary browsing task when peripheral displays are present on the screen.

The average time taken by participants in the Control group (that did not use peripheral displays) was significantly lower than the average time in any of the other test groups. This conforms to the results of both Maglio et al. [5] and Czerwinski et al. [1,2], but contradicts the results of McCrickard et al. [6]. This is interesting because the browsing task used in this experiment is the **exact** browsing task used by McCrickard et al. One reason for the difference in results could be the participant type. It is quite possible that the participants used in this experiment simply could not handle the extra awareness tasks as well as those in the other study. The may be less familiar with the computer, or they may have less experience with peripheral tools. Note that there was no significant difference between the Scale and Fade/Ticker groups, the Scale and Combo groups, or between the Fade/Ticker and Combo groups; which leads to some more interesting ideas.

6.2 There was no significant difference in the performance on a primary task whether using a graphical or a textual peripheral display

The average browse time for the Fade group was not significantly different from that of the Scale group or the Combo Group. Although the awareness task involved in the Fade and Combo groups seems more difficult (reading information, reacting to two different pieces of information, and answering questions about it later) compared to the Scale group (recognize that the indicator fell below the specified level), there was no significant difference between these groups in browsing time on the primary task. This leads us to believe that any display in the periphery affects performance on a primary browsing task. Further research, as described later, would be necessary to corroborate this claim.

6.3 There was no significant difference in the performance on a primary task when using a single peripheral display or multiple peripheral displays.

The average browse times showed no appreciable change when one or two peripheral displays were used. There did not even seem to be a trend towards that result, as the average browse time for the Fade/Ticker group was higher than that of the Combo group. This interesting anomaly is discussed below. We cannot say that this result will generalize to N peripheral displays. Increasing the number of peripherals displays to more than two may or may not affect task performance. This is something that has to be researched and studied so nothing conclusive can be said about the effect of more than two peripheral displays on task performance.

6.4 There was a trend but no significant difference in the number of tasks completed whether using single or multiple peripheral displays.

For the Scale group there were a total of six auxiliary awareness tasks (one per round), 12 in the Fade/Ticker group (two per round) and 18 in the Combo group (three per round). The task completion rate was equal for the Scale and Fade/Ticker groups and was a little less (but not significantly less) for the Combo group. Interestingly, even though the differences were not statistically significant, the actual awareness task completion rate could possibly suggest a trend that increasing the number of awareness tasks may decrease task completion rate even though it does not affect performance on a primary task. The low number of total subjects may have contributed to the lack of significance, or the differences in percentages may have been caused by a handful of users who forgot to complete a peripheral task because they misinterpreted the instructions, were too busy doing the other tasks, or just completely forgot to complete a task.

6.5 The Fade/Ticker group had a slightly higher average browse time than the Combo group.

This was an interesting anomaly that was observed during the evaluation. The Fade/Ticker group had a slightly higher browse time average than the Combo group task. The difference was not statistically significant, but it is interesting to speculate what might have accounted for this difference. Paired with the task completion rates for these groups, a logical explanation for this difference is that the Combo group did not complete as many of the awareness tasks and as a result, the average browse time for that group was lower than the Fade/Ticker group's.

7. Future Work

One aspect of the experiment that could have interesting results is trying to isolate the types of displays (e.g. auditory, graphical, textual) and studying the effects of various combinations of these displays. For example, we could repeat the experiment with a fade only, a ticker only, or a fade and a ticker together or any such combination of peripheral displays that are primarily textual. Another study could use the same idea as above, using graphical displays instead of the textual ones. This will help us in understanding the effects of increasing the number of peripheral displays.

Introducing color into the graphical displays would help in evaluating the effectiveness of the displays in communicating information. The Maglio work used a colored peripheral display, but they did not speculate as to its effects. For example, in future experiments, introducing a color change from green to red in the scale display to

indicate the scale has reached the specified levels might simplify the associated awareness task. It may be very easy for the users to know when to click the button and they would not have to check the status bar at regular intervals. This may increase the rate of task completion or decrease reaction times.

Further studies could be done on the effects of changing the placement of the different displays on the screen. The different displays were placed in the screen in a default fashion that was considered most appropriate as far as the user standpoint is concerned. Many users actually changed the position of the displays to other positions on the screen that they thought was most convenient to them. Would the placement of the different displays on the screen actually affect performance? What was the reasoning behind some users changing their positions?

8. Conclusion

This study was conducted with the primary intent of showing the effectiveness and distraction in communicating information using peripheral displays. The outcome demonstrated a difference in completion time for the primary task when a peripheral display is introduced. The difference indicates that the user's attention is drawn temporarily to the peripheral display. We found no significant difference between the graphical and textual displays in regards to user performance. This result illustrates that the manner in which the information is presented, whether through a graphical or textual display, does not affect the participant's ability to retrieve the information. This seems like a strong result but it must be understood that in this experiment the associated tasks and information for each display type were different. This could be the reason for our results. Further work in the area is required to validate this claim.

Finally, the results showed no significant difference in performance on a primary task when the number of peripheral displays is increased from one to two, though a trend suggests that the number of peripheral tasks that are completed may decrease with added peripheral displays. This indicates that participants were not additionally distracted by more peripheral displays, but instead may have chosen to ignore the additional information.

The results of this study are an exciting early step in understanding the impact of peripheral displays, and could have far-reaching implications. As information becomes more widely available and as computers become more ubiquitous, we expect peripheral displays will be introduced into many aspects of everyday life in order to effectively convey useful information to a variety of recipients. It is essential that we understand how this can be done effectively and with minimal distraction.

9. References

1. Czerwinski, M., Cutrell, E. & Horvitz, E. (2000). "Instant Messaging and Interruption: Influence of Task Type on Performance." In *Proceedings of the Annual Conference of the Computer Human Interaction Special Interest Group of the Ergonomics Society of Australia (OzCHI 2000)*, December 2000.

2. Czerwinski, M., Cutrell, E. & Horvitz, E. (2000). Instant Messaging: Effects of Relevance and Time, In S. Turner, P. Turner (Eds), *People and Computers XIV: Proceedings of HCI 2000*, Vol. 2, British Computer Society, p. 71-76.

3. Heiner, J., Hudson, S., & Tanaka, H. "The Information Percolator". In *Proceedings of the ACM Conference on User Interface Software and Technology (UIST 1999), November* 1999.

4. Hudson, S. E. & Smith, I. "Electronic Mail Previews Using Non-Speech Audio." In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 1996),* April 1996.

5. Maglio, P. P. & C. S. Campbell. "Tradeoffs in displaying peripheral information." In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI 2000)*, April 2000.

6. McCrickard, D. S., J. T. Stasko, & R. Catrambone. "Evaluating Animation as a Mechanism for Maintaining Peripheral Awareness." In *IFIP Conference on Human-Computer Interaction (INTERACT 2001),* June 2001.

7. Mynatt, E., Back, M. Want, R. Baer, M. and Ellis, J. B. "Designing Audio Aura." In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 1998),* April 1998.

8. Weiser, M. & J. S. Brown. "Designing Calm Technology." *PowerGrid Journal*, volume 1 number 1, July 1996.