Evaluating Graphical vs. Textual Secondary Displays for Information Notification

Jacob Somervell, C. M. Chewar, D. Scott McCrickard

Virginia Polytechnic Institute and State University
Computer Science Department, 660 McBryde Hall
Blacksburg, VA 24061

Abstract- This paper reports the findings of a human-computer interaction (HCI) experiment, conducted to determine whether graphical or textual representations of a simulated load monitor are more effective at communicating notification information in a secondary display. We establish guidelines for design tradeoffs based on significant differences in display facilitation of information monitoring, awareness, and introduction of distraction. These findings result from an experiment in which subjects browsed through information pages searching for answers to questions while simultaneously monitoring information in the load monitor. This research is critical in developing a framework for secondary display evaluation that should guide the design and use of visual notification systems requiring a division of user attention.

1. Introduction

Information is everywhere. It is invading our desktops, vehicles, and homes in the forms of instant messages, emails, phone calls, and commercials. Keeping tabs on this plethora of information could easily become a full time occupation. Unfortunately, most people do not have the time to invest in actively seeking information and understanding what they find. As we strive to improve interface design through the study of human-computer interaction (HCI), we are posed with the problem of understanding how to optimally communicate information, or introduce notification of changes, with a computer display.

Within the HCI domain, the intersection of *notification systems* and *information visualization* specifically addresses *secondary displays,* investigating methods for evaluating, designing, implementing, and using graphical representations that capture and reflect important aspects of continuously changing information [2,9]. Information visualizations can enable users to quickly assimilate large amounts of data, and empirical evaluation has led to improved designs over time [3]. However, the evaluation of information visualizations has focused almost exclusively on situations in which users explore the information in a visualization as their only task. In reality, people are often interested in staying informed about constantly changing information, and may desire visualizations for intermittently monitoring a continuously updating information channel. Today, as information from chat tools, Web alert systems, stock trackers, score tickers, geological activity, ocean currents, schedules, and other sources affect desktop computer processing activities, it is important to understand how best to visually communicate important notifications in an effective manner, while accommodating other user task requirements. Secondary display interface evaluation seeks to fill this growing research need.

In this spirit, this paper evaluates four types of information visualizations that could be used in dual-task situations (where user attention is shared between tasks) for supporting secondary notification monitoring tasks. Several characteristics about dual-task systems must be understood to appreciate research and evaluation methods. Generally, a person's attention will be focused on some primary task, but at times it may be necessary to divert partial attention to a secondary task that involves gathering information from a separate portion of the display. This may occur through peripheral vision or shifts in visual focus, but the primary focus of attention is normally expected to remain on the primary task. Hence, only limited attention can be devoted to the secondary display. For example, a student may want to work on a collaborative assignment while watching for chat messages from his colleagues, or an investment professional may want to monitor stock prices while sending email to her clients, or the driver of a vehicle may want to look at map directions while driving. Like other areas within interface evaluation and design, we suspect there are no blanket answers to dual-task design challenges, especially since few dual-task systems have common usage scenarios and requirements.

However, by limiting our purview to graphical and textual encodings at two different update speeds, we seek greater understanding toward the comparative effectiveness for these staple display types at simultaneously filling design goals for notification and dual-task systems. Implications for display design selection and use are founded on the questions we investigate in this work:

- **Which display type best facilitates information monitoring?** That is, we want to determine whether using a graphical representation allows for faster recognition of specific states in the information, as opposed to a textual representation.
- **Which display type best facilitates information communication?** That is, we want to determine whether graphical or textual representations at different update rates are better for promoting awareness, or understanding the data.
- **Which display types introduce distraction to the primary task?** That is, we want to determine whether any of these display types will degrade primary task performance.

Before discussing our experimental methodology, we discuss other research in this field. The experimental methodology section describes our setup and metrics. Experimental findings are presented in the results section and summarized in the discussion section.

2. Related Work

Numerous recent studies investigate various aspects of and techniques for effective secondary display design for notification in dual-task situations. Several studies evaluate or compare other specific information encoding types according to at least one of our research questions. While some earlier studies compare different forms of text animation, more recent research contrasts text with graphics or investigates effectiveness of secondary display graphical properties.

Information monitoring has been investigated in several studies. By evaluating various text-based tickers as secondary displays in a dual-task environment, Maglio and Campbell found that there was no difference in how well participants could recognize headlines in the various types of tickers used [5]. Since animation is now being widely used in display design, especially in web pages and advertisements on the Internet, Bartram's study of the effectiveness of animation as a communication device is particularly interesting. They found animation is more effective for monitoring than both color and shape, when used in the periphery [1]. McCrickard et al. had subjects perform a browsing task and simultaneously monitor various information sources in text displays [6]. The displays employed various animation techniques, including fading, scrolling, and immediate updates. Their findings indicate fade displays are better for facilitating monitoring.

The relative effectiveness of notification display support for conveying understanding, or awareness of information, has also been investigated. McCrickard's study also indicated scrolling tickers were better for understanding secondary task information [6]. Another study recently completed by Tessendorf, et al. [10] found images in a user's focus allow more insight than images in a secondary display, and users always gained the most insight from positional representations.

Our final research question concerns the introduction of distraction from the primary task. Nearly all work related to dual-task research investigates this causal relationship. Most researchers find that the introduction of a secondary task negatively impacts performance on a primary task [4, 5, 7, 10]. However, McCrickard's study found no significant introduction of primary task distraction [6]. Specifically, Czerwinski et al. observed the distraction effects of instant messaging on database search tasks [4]. Maglio and Campbell's text-based ticker study notes that when the ticker was present, there was a negative impact on editing performance [5]. Somervell et al. performed an extension to the McCrickard experiment [7] to include graphical information displays. Again, they found that the presence of the information displays negatively impacted primary task performance.

Other aspects of dual-task research develop design tradeoffs based on variation in system design goals. McCrickard discovered performance tradeoffs among display types that correlate with users goals of remembering or recognizing the presence of information [6]. The Tessendorf study indicated another design tradeoff—at low levels of primary task degradation color is a better information representation than area; however, at high levels of degradation, area is better than color [10].

3. Evaluation Methodology

This section provides a description of the empirical process used to learn about relative secondary display effectiveness of two information encodings, presented at two update rates. The discussion of experimental setup describes the test program used, independent variables, test population and conditions. A detailed explanation of experimental metrics follows, to include definition, discussion of relevance, acquisition of metrics within this experiment, and examples of practical use.

3.1 Experimental Setup

In order to investigate the utility of graphical and textual displays as secondary information tools, a dual-task environment was created in which subjects were asked to perform a simple browsing task (primary task) while simultaneously monitoring information about a simulated computer load (secondary task). The goal was to create an environment that served as a model of a typical activity. People often browse the Internet looking for information, either for work or pleasure, so the primary task used in this experiment models this activity. As for the information-monitoring task, we chose a computer load as the information source, both because it is a familiar concept and because it could be depicted both graphically and textually.

The browsing task involved using a simple web browser, much like Internet Explorer or Netscape Navigator, to traverse information pages looking for the answer to a specific question. A typical question would be something like, "How many daily tours of the Martin Luther King, Jr. birthplace are there?" Each of the eight rounds in the experiment consisted of a unique set of four questions and corresponding information pages. The answer to each question was always a numerical answer to avoid typing errors. Participants advanced to the next round only after correctly answering all four questions.

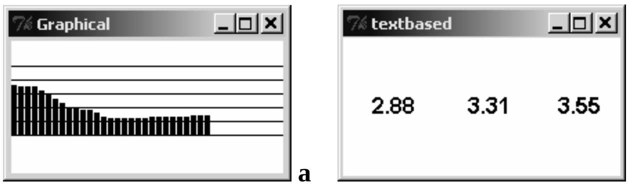
The secondary task in this experiment required monitoring of a simulated computer load, represented either graphically or textually. The graphical visualization consisted of a moving vertical bar graph with each bar representing the load for one second (see Figure 1a). The load was indicated by the length of the bar, as it was imposed on a scale from 0 to 5, with the top of the bar representing the load level. The display contained 30 such bars, hence providing information about the load over the last 30 seconds. When updated after the first 30 seconds, the oldest bar moved off to the left and the new bar was shown on the right of the display.

The text-based visualization consisted of three numbers, tabbed horizontally (see Figure 1b). The leftmost number represented the current load. The center number represented the average load over the last five seconds. The rightmost number represented the average load over the last 30 seconds. Hence, the text-based display presented the same information as the graphical display. When updated, all three numbers changed to indicate the new information.

There are two independent variables used in this experiment: encoding and update rate of the secondary displays. As mentioned, the display type was either graphical or textual with a fast (300 ms) or slow (3000 ms) update rate. Sixty-three undergraduate computer science students participated in the experiment for course credit. The experimental software was written using Tcl/Tk 8.3 and ran on Windows 98 PC's in a closed lab environment. Five test groups were used to cover the four experimental conditions plus one control group that only performed the primary browsing task. Twelve or thirteen participants formed each test group, and completed eight rounds each. Data on round completion times, state recognition times, and multiple-choice answers to the after-round questions (as discussed below) were collected electronically.

3.2 Metrics

Our objectives were to determine which display (if any) is better for facilitating monitoring, communicating load information, and minimizing distraction to the primary browsing task. This section supplies a detailed look at our metrics: *monitoring latency* and *monitoring response*, *information awareness*, and *distraction to the primary task*. This includes a brief discussion of each metric's relevance to possible design considerations, providing examples of practical use and explaining how measurements of it are obtained in this experiment.



a b

Figure 1: Secondary displays used in the experiment. a) Graphical representation of the computer load. The leftmost bar is the load 30 seconds ago, the rightmost bar is the current load. b) Textual representation. The left number is the current load, middle is the average load over the last five seconds, and the right number is the average load over the last 30 seconds.

Monitoring can be analyzed in terms of latency and indication of recognition with a response, but it is always associated with recognizing presence, or a certain state of information. Desired information state is specified in the experiment by a single question visible for the duration of the round. The tasks the participants were to complete depended directly on the information in the displays. Two types of questions were used in the information-monitoring task: recognizing specific levels and recognizing change in direction. For example, participants might be asked to indicate when the load falls below three or when the load starts to decrease. They indicate that they have seen the desired state by clicking a button on the screen.

Monitoring latency describes the difference in time between a display depicting the given information state (state activation) and a user acknowledging recognition of that state. For example, given that participants are asked to identify when the load starts to decrease, if a participant recognized that the load began decreasing seven seconds after it actually reached that target state, then the monitoring latency would be seven seconds. This is a critical metric for understanding communication effectiveness—especially for unobtrusive displays. If monitoring latency is negligibly small, then we can have some confidence that a user's information monitoring is continuously maintained. Monitoring latency as a design feature becomes important if the information displayed in the secondary task is time-critical; someone concerned with instantaneous stock prices, medical information, network traffic from a system administration perspective, or information supporting computer-supported cooperative work would likely be concerned with minimizing monitoring latency.

Monitoring response is a binary condition—a participant either demonstrates monitoring of information or fails to demonstrate monitoring. A system displaying non time-critical information may not have a need for low or optimal monitoring latency, yet may still have a requirement for monitoring response. Other time-critical information displays may also be best evaluated by monitoring response if the unit of time by which criticality is defined is on the order of several minutes or hours, rather than seconds.

Information awareness describes how well a user has gleaned an overall understanding about the data. This metric is comprised of an average correctness score of six multiple-choice questions about the load at the end of each round. Questions tested recognition and understanding of minimum/maximum values, overall trends in the data (increasing vs. decreasing), overall averages, comparisons between total time above certain levels as compared to below other levels, time spent above/below specific levels, and overall variability. A participant's ability to answer these questions correctly relies on maintaining monitoring of display action (perhaps by glancing at the secondary display for two or three seconds every ten to fifteen seconds), but also requires higher level processing of the data. In order to know whether the overall system load trend was increasing or decreasing or what portion of the time it was over a given level, a participant had to create some memory of the overall dataset. We assert that a characteristic of a display facilitating effective information awareness is the ability to assist memory creation. Quite possibly there are many situations in which overall memory or understanding of data displayed in a secondary window would not be critical or even desired. Often, some level of basic monitoring may suffice. However, other design requirements and usage scenarios may need to address understanding and analysis of secondary data, making this metric key in a comparative evaluation process.

When a system explicitly requires a user to divide his attention between two or more unrelated information processing tasks rather than focusing on a single primary task, it seems impossible not to introduce primary task distraction. In other words, unless a user is completely ignoring the secondary task, some efficiency must be lost in the primary task performance. In this experiment, primary task efficiency is captured by round completion time. The differences in round completion times between the control group and the groups with secondary displays can be said to capture the distraction caused by the secondary task, since no other aspect of the experiment varied. Certainly, other examples of primary tasks would have a wide range of efficiency measurements, and for many primary tasks this measurement may be trivial. However, there are many conceivable examples that extend far beyond the desktop computer where primary task distraction must be kept minimal—displays supporting vehicular, medical, and military operations certainly would value minimal introduction of distraction. The following section reports the statistical findings from this experiment.

4. Results

This study empirically establishes several relationships between our five test conditions across three important design objectives: facilitation of information monitoring, awareness of information, and introduction of primary task distraction. We found that adding a secondary notification task to the browsing task has a significantly negative impact on task completion times ($z(103)=2.07$, $p<0.05$). Therefore, adding a secondary display should be done with judicious consideration to the value added in support of user goals. Furthermore, our findings indicate no single encoding or update rate offers optimal performance for all three objectives. Rather, the four display types (abbreviated as *slow-text*, *fast-text*, *slow-graph*, and *fast-graph*) present different combinations of strengths and weaknesses, forming a collection of design tradeoffs according to facilitation of information monitoring, awareness of information, and introduction of primary task distraction. All difference tests were performed as sample confidence interval comparisons to the population mean. Since our comparison sample sizes were greater than 30 and data distribution was approximately normal, we used z-scores to assess statistical significance. The following sections detail the findings under each design objective.

4.1 Facilitation of Information Monitoring

The first objective of our experiment was to determine whether slow-text, fast-text, slow-graph, or fast-graph representations of information in a secondary task display allow for quicker recognition of specific information states. The data allow analysis toward comparing the four display types within two distinct design considerations: *monitoring latency* (difference between state activation and participant state-recognition times) and *monitoring response rate* (percentage of participants who identified correct state).

Monitoring latency comparisons resulted in no reportable experiment-wide display type differences. Several rounds did have significant differences among display types, however these results were inexplicably contradictory between rounds and apparently were not strong enough to influence the overall experiment. However, by analyzing sample differences between display types at cumulative round completion times, many significant differences are apparent (see Figure 2).

For example, slow-text displays yield significantly lower monitoring latency times when compared to at least one (and sometimes all three) other display types from round completions at 250 up to 450 sec. For round completion at 275 up to 300 sec, fast-text displays create significantly higher monitoring latency than all other types. Although fast-graph displays appear to require higher monitoring latency at round completions accumulations of 350 sec and higher, the difference is only significant with slow-text displays.

Monitoring response rate comparisons are based on a count of participants for each round that achieved state transition monitoring. The experimental setup should result in equal portions among the display types in a given round, although round-by-round comparisons would lack validity. Figure 3 shows this data, although clearly few rounds exhibit equal portions by display type. Between-group variation prevents any overall statistically conclusive

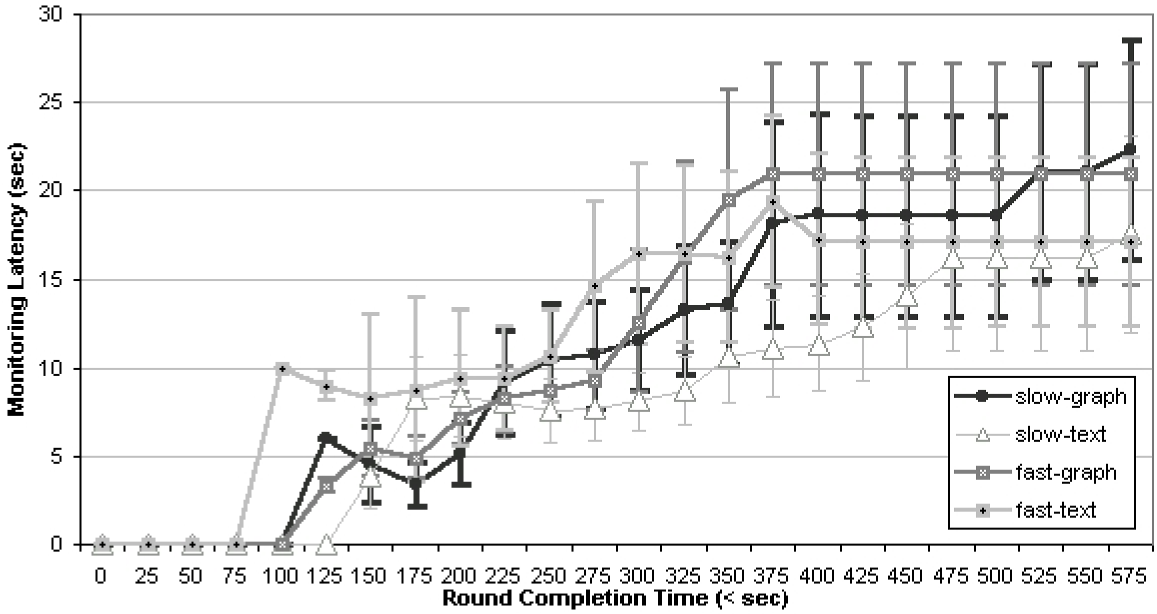


Figure 2: The mean monitoring latency (indicated by symbol) and 95% confidence intervals for all four displays types are shown for each cumulating group of subjects within round completion times. Sample means outside confidence intervals of other samples indicate statistical significance.

finding about differences in monitoring effectiveness. However, the slow-text groups had the highest mean percentage of subjects achieving monitoring states *(mean = 66, σ=29)* and the fast-graph group had the lowest *(mean=55, σ =33).*

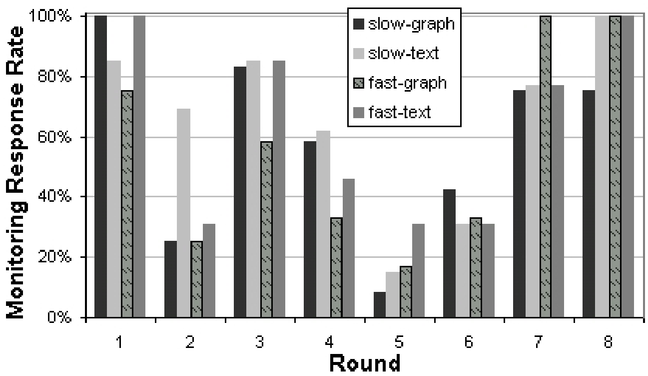


Figure 3: Percentages of participants in each group that responded to information state-change through monitoring each of the four secondary display types.

4.2 Awareness of Information

The second objective of the experiment was to determine which display type allows a user to achieve better understanding of trends, averages, and overall meaning from the data. This understanding was tested with six multiple-choice questions asked after each round. The scores of these questions were totaled for each round, prior to comparisons between display types. Two overall findings are significant: compared with the population mean, 1) fast-text displays resulted in lower correctness scores (z(103)=2.12, $p<0.05$) and 2) fast-graph displays yielded higher correctness scores (z(95)=2.50, $p<0.05$). Figure 4 shows how the four display types vary in promoting awareness of information according to round completion times. Most remarkable is the consistent ordering of the display types maintained for cumulated round completion times after 225 sec. Low variance within that band allows further comparisons between display types to yield statistically conclusive results. For example, slow-text displays almost always yield higher correctness than fast-text, and fast-graphs outperform slow-graphs in nearly all cases.

4.3 Introduction of Distraction

A final evaluation criterion for effectiveness of secondary task displays is the degree of distraction it introduces upon the primary task, discernable as a negative impact on primary task performance. In this experiment, primary task performance is defined as the time required to correctly answer four questions based on a browsing task—measured by the round completion time. The control group round-completion times provide a benchmark for performance without distraction. Two significant findings were apparent throughout the experiment: compared to the population mean, 1) the control group completed rounds significantly faster (z(103)=2.07, $p<0.05$) and 2) the slow-graph display group completed significantly slower (z(95)=2.01, $p<0.05$). Pairwise comparisons of display type samples also show

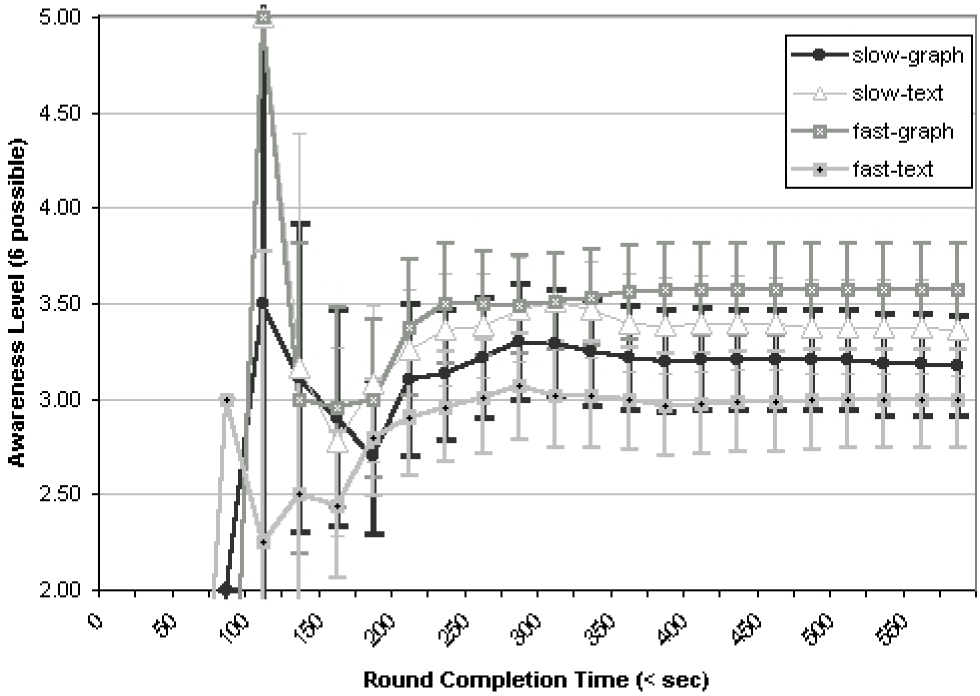


Figure 4: The mean awareness scores (indicated by symbol) and 95% confidence intervals for all four displays types are shown for each cumulating group of participants within round completion times.

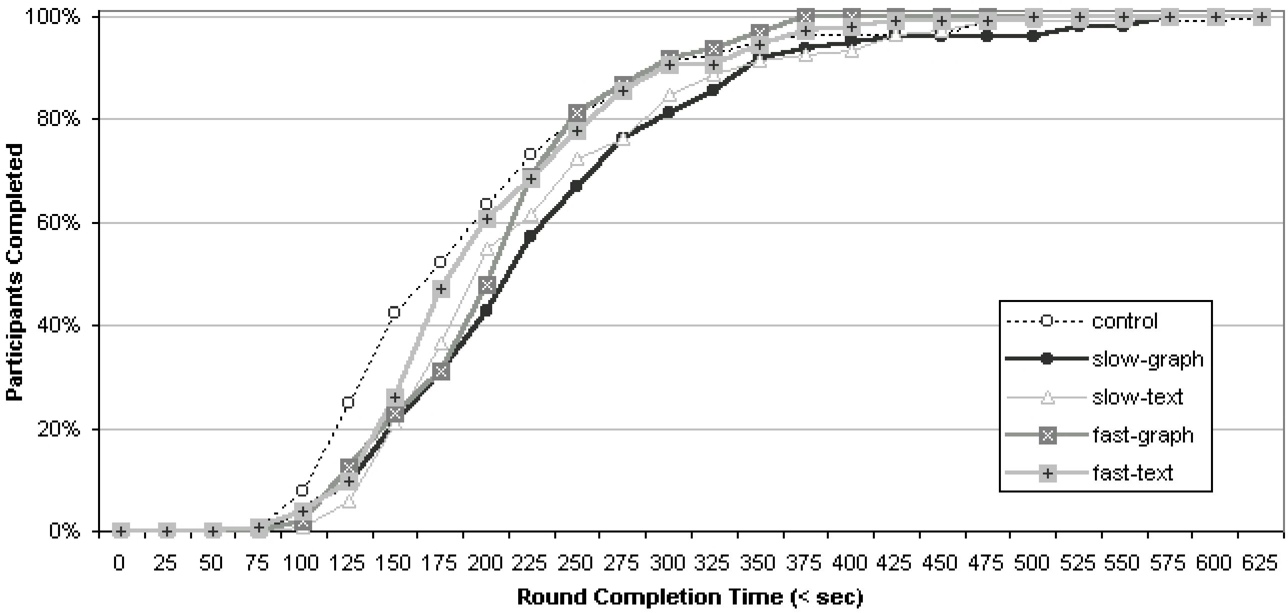


Figure 5: Cumulating groups of participant round completion times according to display type or control group (none).

that the fast-graph and fast-text displays had faster completion times than the slow-graph displays (z(95)=2.05, p<0.05), (z(95)= 2.30, p<0.05). Figure 5 shows the accumulation of participants according to round completion times. Note that at 150 sec, approximately 43 percent of the control group had completed their rounds, but it took slow-graph users another 50 sec to reach the 43 percent completion level. Such a 50-second performance difference in the first 200 seconds indicates this significant difference is quite possibly noticeable and relevant as well. Figure 6 represents the same information shown in Figure 5, but as trendlines (regressed, sixth order) of twenty-five second time bands, rather than depicting cumulated data. This allows easy contrast of distribution skew and span. For instance, slow-graph users tend to have greater dispersion across the range of round completion times than other display types.

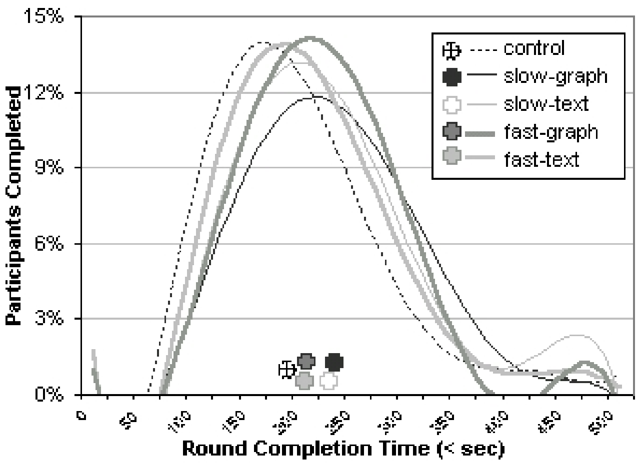


Figure 6: Trendline histogram representing bands of population round completion times according to display type. Mean round completion times are indicated near the top.

5. Discussion

Having analyzed the performance data associated with the four different display types and the control group, this section provides a summary of experimental findings, comparing display types by design objective and their relative performance strengths and weaknesses. If system design requirements specify features such as monitoring, information awareness, or minimal primary task distraction, results of this experiment may be used to make recommendations about displays types most suitable for secondary information notification.

To summarize experimental findings, we provide positive and negative recommendations for display type selection according to design objective (Table 1). Of course, several conflicting or non-measured objectives may exist, indicating a need for criteria prioritization or further research. Similar summary information is presented in Table 2, although this is arranged according to display type. Both tables only contain design tradeoffs that can be identified based on the results of this experiment.

Several remarks can be made about the experimental results. If a display can be characterized by low monitoring latency and low distraction to the primary task, this impressive set of design features indicates support for an effective split of user attention, seemingly quite useful for dual-task notification system design. Unfortunately, these design features tend to become tradeoffs and have not been found as concomitant strengths within a single display type.

In overall experiment results, monitoring response rates were surprisingly low. Failure to respond in recognition of a specific information state can be attributed to several causes: the information encoding failed to facilitate comprehension necessary to recognize the state change, participants ignored the secondary task, or participants completed the primary task (thus ending the round) prior to reaching the state change. However, even after a thorough contrast of Figure 3 and mean round completion times by display type, no pattern seems to be apparent to suggest one of these causes may have prevailed. It is disappointing that none of the display types were distinguishably superior or inferior under this metric, since it is an important aspect of monitoring facilitation.

6. Conclusions

This work has focused on evaluating the effectiveness of graphical and textual information representations in secondary displays designed for notification. A dual-task setup was used to determine which display was better at facilitation of information monitoring and awareness, as well as preventing distraction from a primary task. No single display type was simultaneously optimal for all three goals, but we found some interesting design tradeoffs (see tables 1 and 2).

It is important to note the limited application of these results. The dual-task setup used in this experiment models a tiny subset of the possible notification monitoring situations. The simple browsing task, while a relevant task, does not demand high attention from the participant; unlike other tasks such as editing or searching. Furthermore, the information-monitoring task used in this experiment is only one of countless tasks people engage in while they are busy doing other tasks. Using strictly numerical information as the source for the secondary task also limits the applicability. For example, using a slow-text display is best for minimizing monitoring latency for our computer load, but a fast-graph display could prove best for electronically tracking wildlife.

The major contribution of this work is the evaluation methodology, to include the metrics, experimental setup, and collection of design tradeoffs. The extensibility of these techniques is paramount to ongoing research into effective designs for notification systems. Other efforts within this research area investigating information density and presence time [8] and effectiveness orderings of design attributes for secondary displays [10] can benefit from this evaluation framework. The design tradeoff findings outlined in this paper help us understand the most expressive methods for displaying numerical information that is used in visual notification systems or secondary displays. Further research should focus on improving generalizability of results, broadening tradeoff tables by including additional display types, and developing dual-task notification systems to user specifications, allowing field testing to validate implementations of our design guidelines.

Table 1. Recommended secondary display types by design objective

	Recommended	**Not Recommended**	**Comments**
Monitoring Latency	Slow-Text	Fast Displays	Significant results limited to groups based on round completion time
Monitoring Response Rate	Slow-Text	Fast-Graph	No significant results, recommendation based on mean performance only
Information Awareness	Fast-Graph	Fast-Text	$p < 0.05$
Minimal Primary Task Distraction	No Secondary Task, or else a Fast Display	Slow-Graph	$p < 0.05$

This table can be used to identify the most suitable display type according to one of the four specific design objectives.

Table 2. Secondary display type strengths and weaknesses

	Strengths	**Weaknesses**
Slow-Text	Monitoring Latency & Response Rate*	Below Average Mean for Primary Task Distraction*
Fast-Text	Minimal Primary Task Distraction	Monitoring Latency & Awareness
Slow-Graph	None	Primary Task Distraction
Fast-Graph	Awareness & Minimal Primary Task Distraction	Monitoring Latency & Response Rate*

The four display types tested in this experiment are compared according to their strengths and weakness in design objective performance. * = $p>0.05$, table entry based on mean only.

7. References

1. Bartram, L., Ware, C., & Calvert, T. Moving Icons: Detection and Distraction. In *Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT 2001)*, Tokyo, Japan. July 2001.

2. Card, S., Mackinlay, J. & Shneiderman, B. *Readings in Information Visualization.* Morgan Kaufman, 1999.

3. Chen, C. & Czerwinski, M. Empirical Evaluation of information visualizations: An Introduction. *International Journal of Human-Computer Studies, 35(3), 2000.*

4. 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 200*0, Vol. 2, British Computer Society, p. 71-76.

5. Maglio, P. & Campbell, C.S. Tradeoffs in Displaying Peripheral Information. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2000)*, April 2000.

6. McCrickard, D.S., Catrambone, R., & Stasko, J. Evaluating Animation in the Periphery as a Mechanism for Maintaining Awareness. In *Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT 2001)*, p. 148-156, Tokyo, Japan, July2001.

7. Somervell, J., Srinivasan, R., Woods, K., Vasniak, O. Measuring Distraction and Awareness Caused by Graphical and Textual Displays in the Periphery, In *Proceedings of the 39th Annual ACM Southeast Conference*, Athens, GA. March 2001.

8. Somervell, J. McCrickard, D.S., North, C. & Shukla, M. An Evaluation of Information Visualization in Attention-Limited Environments. *In Joint Eurographics IEEE TCVG Symposium on Visualization (VISSYM 2002).*

9. Spence, R. *Information Visualization.* Addison-Wesley, 2001.

10. Tessendorf, D., Chewar, C. M., Ndiwalana, A., Pryor, J. McCrickard, D. S., and North, C. An ordering of secondary task display attributes, *Submitted to Conference Companion of the ACM Conference on Human Factors in Computing Systems (CHI 2002)*, April 2002.