Usability Testing of Notification Interfaces: Are We Focused on the Best Metrics?

John E. Booker, C. M. Chewar, and D. Scott McCrickard
Center for Human-Computer Interaction and Department of Computer Science
Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0106 USA
540-231-7409
jobooker@vt.edu, {cchewar, mccricks}@cs.vt.edu

ABSTRACT
Notification interfaces that continuously present peripheral information have received increasing interest within the HCI community, especially those supporting awareness of others' activities. While recent empirical studies have focused on information design aspects of peripheral displays, there have been few reported studies that comparatively evaluate actual systems. To this end, this article describes our efforts in comparing three interfaces that inform a remote user about activities within a given setting. Our data allow conclusions about comparative interface usability and preference, and provide an indication about metrics that are valuable to focus on in evaluations for these types of interfaces. In particular, we find that quantitative, performance related metrics, such as the correctness of notification interpretation and interruption to a primary task, are much less conclusive for fully implemented peripheral interfaces than qualitative judgments based on the usage experience.

Categories and Subject Descriptors
H.1.2 [**Models and Principles**]: User/Machine Systems - *Human Factors*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Evaluation/Methodology*

General Terms
Design, Experimentation, Human Factors

Keywords
notification systems, peripheral displays, monitoring, empirical evaluation

1. INTRODUCTION
People often want to monitor activities of others without maintaining a physical presence or imposing upon privacy. For instance, parents may want brief liberation from their young children's play activities, although concerns for safety of children, even in the next room, may necessitate frequent inspections. While privacy may not be a necessary tradeoff in monitoring one's children, supervisors are often uncomfortable with "looking over their employees' shoulders," although they maintain interest in characteristics of their work activities and patterns. While these situations reflect typical supervisory functions, this information need can also be motivated by teamwork concerns in a distributed collaboration effort—remote group members are often interested in activities of co-located members.

Advances in computer vision, live video capture and transmission, and networking technologies have made real-time, remote scene monitoring feasible and inexpensive from an implementation perspective. Despite having needs to monitor activities of employees or team members, many people are unwilling to use these systems for a variety of reasons. Often, this inhibition involves an uncomfortable feeling associated with watching others—a consequence of invading the social expectation of privacy. However, since these systems would usually be monitored while users are engaged in other tasks, there may be problems associated with the undesirable amount of attention required to observe and track remote events.

This paper focuses on the issues resulting from attempting to represent clear information about a remote scene while maintaining the privacy of others and not interrupting the user. Both of these problems are certainly within the interest of the human computer interaction community. In this paper we also describe the difficulty in evaluating such a system with basic research test methodology. Encouraged by the recent progress that has been made toward supporting group activities without encroaching on privacy, as well as designing and evaluating notification systems, we have a general goal of developing and assessing new approaches to the scene monitoring information need.

2. RELATED WORK
Research in computer supported collaborative work (CSCW) has made great strides in understanding how to portray group member activities and preserve privacy. As an alternative to direct audio or video, Dourish and Bly explored methods of providing background awareness of work groups with their Portholes clients [8], although the interfaces were photographic images that did not account for privacy concerns. The tradeoff between supporting awareness of scene details and preserving privacy was explicitly recognized in Hudson and Smith's 1996 work, in which they responded by introducing three privacy preserving techniques that provide the benefits of informal serendipitous interaction to distributed work groups: the shadow-view video feed, an "open-microphone" shared audio technique that removes all intelligible words, and a dynamic group-photo that indicates presence or absences of co-workers. Other work has focused on refining video techniques to maximize privacy, assessing the comparative impact of blur or pixilization at various fidelities on awareness and again noting the tradeoff awareness and privacy [3]. Greenberg and Kuzuoka [10] address this recognized tradeoff,

Conference'04, Month 1–2, 2004, City, State, Country.

providing a very innovative approach with their "digital but physical surrogates"—tangible, often comical objects that represent individuals, achieving various levels of peripheral information perception. However, all of these methods seem to be fairly interruptive to ongoing tasks and do not provide any sense of context or history (i.e., how the current state is different from several minutes ago).

The AROMA project extends the application domain for representing remote activity and presence from workplace to living space with an approach that captures activity data, abstracts and synthesizes them into streams, and displays the information with ubiquitous media [14]. The authors provide a compelling argument about the importance for history and memory support, as well as a sound architecture for a generic system. The recent notion of social translucence and its prototypical interfaces [9] also address the awareness-privacy issue in an exciting way, using simple abstractions to chart activities of group members. Most importantly, the social proxies introduced by these researchers are embedded within a larger groupware application, an implicit acknowledgement of the user's tendency to expect this type of information as a secondary information processing task.

The emerging HCI area of notification systems research specifically investigates the design and evaluation of interfaces that are typically used in divided attention situations with a low portion of focused attention [13]. Like the digital surrogates and social proxies, several notification ideas have shown promise in providing awareness of remote persons of interest. The informative art (infoArt) interfaces are a novel and aesthetically pleasing approach that can convey many dimensions of information, as well as historical context [15]. Other systems, although not intentionally designed as notification systems, show renewed potential for the use of face-like animation. Jeremiah [2], in particular, drew unprecedented interest by onlookers, as it responded to abstracted scene information obtained through its vision system with changes in emotion and gaze orientation.

While the work in the CSCW field inspires confidence that successful interfaces can be designed to support monitoring of activities within a remote location, we are uncertain how usability of a notification system implementation can be optimally assessed. Very few usability studies of fully implemented notification systems appear in literature, and often only include analysis of user survey responses based on system experience, e.g. [4]. We are hopeful that system logged and task embedded usability performance metrics for assessing dual-task situations, such as those that were indispensable in basic research [1, 6, 7, 12], will be influential in comparing various notification displays that supply scene activity information. preference-related survey questions.

2.1 Project Objective

To assess the notification systems usability test methodology for fully implemented interfaces and determine the usability of a notification system that applied the guidelines from awareness-privacy literature, we designed a vision-based system that senses the presence of people and would deliver remote scene characteristics to a user as a notification system. The work here describes the design and evaluation of three interface prototypes that employ the use of both preference and performance sensitive metrics.

3. INFORMATION & INTERFACES

In specifying the criteria for the interfaces, we wanted to avoid intruding upon privacy while still leading the users to correct inferences about the scene. We used these criteria to select which scene characteristics we would represent, as well as how they would be represented. We identified six parameters of group behavior and ordered them by their importance as scene characteristics. These parameters were implemented as states that would be conveyed with the interfaces. The six states were:

- population—continuous variable showing the number of people present within the scene (up to ten people total at a granularity of two), as determined by the vision system
- movement—three discrete levels indicating the activity levels of the majority of people—no movement at all, quiet movement (fidgeting, writing, or tapping a pencil), or active movement (occupants moving around the room)
- location—representing the general position of room occupants as either all standing up, most standing up, most sitting down, or everyone sitting down
- familiarity—determined by face recognition, representing the ratio of strangers to known occupants present with three levels—no strangers, some strangers and some familiar people, or only strangers; additionally, whenever strangers entered an otherwise empty room, the interfaces alert the user
- collaborative work—three levels conveying whether all, some, or no occupants were working together; determined by the angles and patterns of face orientation and body proximity
- time—relating the amount of time that had passed since a state change within the scene, letting the user know the newness of the displayed state

The most important states were mapped to the most visible display attributes. Since the interfaces were designed to be secondary displays, they appeared in the lower right hand corner of the screen and used an area of 150x150 pixels on a desktop in 1074x768 mode. As secondary displays, we did not want any of the interfaces to be interruptive to other tasks that the user was engaged in, so all of the animations were made as smooth as possible. The general design objective was to create interfaces that could be easily and quickly understood when looked at, but would not draw the user's attention away from other tasks as the displays transitioned (the exception to this was the unfamiliarity alert). To accomplish his, we designed and tested three interface prototypes: a face depiction (Smiling George), an infoArt option (Spinning Cube), and a simple bar chart (Activity Graph) (see Figure 1).

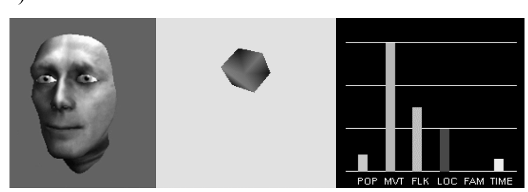

Figure 1: George, the Cube, and the Graph

3.1 Smiling George

Using the DECface[1] platform, we created a facial animation (referred to as George) that allowed us to map the five states to

[1] by Keith Waters, Available at: http://www.crl.research.digital.com/publications/books/waters/Appendix1/appendix1.htm

individual display attributes in a highly metaphoric way. Since George could express emotion, it was excellent for our purposes. George was designed to respond to the scene as if he were a direct observer. Therefore, population was represented by the degree of smile—the more, the merrier. Movement and location of students were represented by the movement of the face within the display window and the level of the face's vertical gaze, respectively. The presence of unfamiliar students was indicated by a red window border (the border was not present if everyone in the room was known). Smiling George indicated the degree of collaborative work by the speed that it shifted its gaze back and forth horizontally—if everyone was working together in one group, then the face stared straight ahead, leveraging the metaphor that George would attempt to look at the different groups working. As new events occurred, the background brightened and then faded to black after about a minute. The brightening of the display was not meant to be an alert, so it happened very smoothly over a short period.

3.2 Spinning Cube

We wanted an aesthetically appealing interface option, so we designed an infoArt cube that spun rhythmically within the display window and changed its appearance based on the environment's status. It would act similar to the face, but would convey information without side effects resulting from possible connotations associated with various facial animations. Population was proportional to the size of the cube. Movement was mapped to the rotation of the cube, while location was represented by the vertical position within the window. The amount of collaborative work was represented by the amount of green hue on the otherwise blue cube. The time elapsed since the last event was represented by the same fading background as used for Smiling George.

3.3 Activity Graph

We designed a bar graph interface to be a simple, low-abstraction interface, thus it did not make use of animation and color changes. The graph consisted of six vertical bars of different colors, with three horizontal background lines for scale references. It was an atypical bar graph, since each bar did not have the same scale. Population has ten values, but movement and familiarity have three discrete values. Thus for the latter two states, the bar was either at zero, at half of the max value, or at the max value. Underneath the x-axis were abbreviated labels for each bar. The familiarity alert was the only event not represented in graph form—when the level of unfamiliarity increased, the background flashed red.

4. USABILITY TESTING

Having designed three interface prototypes we conducted user testing to draw conclusions about the notification systems test methodology and compare the different visualizations methods.

4.1 Hypotheses

We were eager to identify which of our interfaces designs had the most potential for continued development. Like any successful notification system, we expect that the different interfaces will have no significant, unwanted interruption effect on the ability of users to perform their primary task—both in task related accuracy and pace, as compared to task performance without the notification task.

1. We expect that differences between interfaces in effects on primary task performance and comprehension of scene-related information will provide the most poignant testing results.
2. However, we anticipate common performance characteristics in specific features-mappings (e.g., use of horizontal motion range or brightening of display background) that are included in multiple interfaces.
3. Finally, we expect minor differences in preference-related survey questions.

4.2 Participants

Participants for this experiment were primarily male computer science majors 18 to 20 years old. A total of 80 students participated. Of these, 11 were considered expert designers and participated in a pilot study which isolated flaws in the interface and helped target areas of interest. While 69 participated in the final version, only 67 were used in the data analysis. None of these participants had any significant exposure to HCI principles, so we consider them to be expert users rather than novice designers. Participation incentive was class credit.

4.3 Procedure

Our lab-based experiment was run on up to nine participants at a time who were paced together through the experiment. Instructions were given both on participants' individual machines and on a large screen display so that users could follow along while an experimenter read aloud. The first set of instructions introduced the students to the experiment and set up the test scenario: Acting as a professor in a remote location, they wished to monitor a lab, but for reasons of privacy could not use a direct video image. They were also instructed in the use of the primary interface—a spreadsheet-like interface for class grade calculation and entry. The participants were to sum the highest three of four quiz grades and enter the total in a textbox—this task was repeated for an indefinitely long series of quiz grades, serving to focus attention away from the interfaces in question.

After the overall introduction, participants began the first of three timed rounds. The order of interface testing was counterbalanced among six groups by a latin square design. Thus we had three different test groups for two versions: one with and one without a primary task. This made for a total of six versions, each to which we assigned between 10 and 13 participants.

Each round started with instructions for the tested interface. The instructions consisted of screenshots of the interface representing the levels of all states, along with textual explanations. Users then moved on to the interface, monitoring the secondary display which was driven by a simple *activity script file*—a different one for each round. As they viewed the scene monitors, they were also calculating grades if their version included the primary task. To compare performance on the primary task across interfaces, we measured the time between grade calculations. This allowed us to determine a *grading rate*, which was the average of the differences between grade entry times. However, this only told us how fast they were computing grades, not how well. Therefore, we considered the correctness of each grade, which we used to calculate the percentage of correct grades, or *grading accuracy*. These two scores allowed us to evaluate the primary task performance. High performance here would indicate that users were able to work uninterrupted by the secondary display.

In addition to testing if the notification interface was interruptive, we also had to test if the interface was informative. We did this upon completion of rounds. Rounds ended when the activity script ran out (after about five minutes), but users that had the primary task were made to believe it ended because they finished all their grades. This encouraged them to expedite their grading and primarily focus on this task. Once done, users' ability to comprehend information in the secondary display was evaluated with a series of five *scene recall* questions that were unique to each round. These questions asked users about the activity of the remote lab's occupants during the preceding round (e.g., *what the largest number of students at any given time?*). A high score here meant that users both saw and correctly interpreted the information provided by the secondary display. If users had a version without the primary task, then users constantly monitored the information with full attention, and thus the score would only be affected by the interface version.

To measure the users' perception of the interfaces' ability to provide functionality, at the end of each round participants were presented with a series of nine *interface claims* designed to identify the interfaces' perceived impact on interruption, comprehension, and reaction. Users agreed or disagreed with these statements according to a seven point scale, where agreement always indicated a positive perception of the interface. While actual interruption and comprehension would be determined by performance metrics, we were also interested in determining user satisfaction with the interface. One might infer that an effective interface would be an appreciated one, but we wanted to find out from the users directly. Thus, these additional questions were needed to assess the total user experience.

When all the participants had answered the scene recall and interface claim questions, the round terminated and a new one started with the next interface. Once all rounds were finished and all interfaces seen, users were asked to choose the best interface for a series of *ability awards* (e.g., *which was easiest to learn*?). Thus, in addition to performing our own tests for significant differences in the interfaces, we could ask the users if they thought there were important differences.

5. RESULTS

After running the experiment and performing the analysis, we organized the results into three sections based on the three hypotheses. We start with data that address the first hypothesis, saving discussion for a later section.

5.1 Overall Performance Metrics

To investigate the first hypothesis, we looked at how well each interface supported the primary task's grading rate and grading accuracy, as well as the secondary display's scene recall questions. The primary task data were collected across the entire five minutes, and for the scene recall data each participant was given a score for how many correct answers he/she provided out of five. This aggregated score is examined in this section.

We first looked at the grading rate, or how fast participants performed the primary task. When the data among the interfaces were compared, we found the averages for the graph, cube, and George were 12.9, 9.8, and 8.7 seconds, respectively. The overall averages of each participant's standard deviations were 9.2, 5.3, and 4.8 seconds. We found no significant differences in grading rates.

Next, we examined the correctness of the primary task—grading accuracy. The averages for the graph, the cube, and George were 96%, 94%, and 96% respectively, with standard deviations of 3%, 6%, and 5%. As with the grading rate, differences in these performance results were not significant.

For the scene recall questions, the overall percentages of correct answers for the versions with the primary task were as follows: graph-47%, cube-37%, and George-40%. The standard deviations were 24%, 28%, and 24% respectively. When a primary task was not present, the scores in order of graph, cube, and George were 56%, 44%, and 51%, with standard deviations of 26%, 27%, and 28%. Differences among both these sets of results were insignificant. Additionally, for any of the interfaces no differences were found when comparing between the participants that were tested on the version with a primary task and the version without.

Thus, we found no significant difference in any of the overall performance metrics.

5.2 Specific Performance Claims

In this section we take a closer look at the scene recall data, broken down into individual questions. There were several cases where scene information was depicted in two different interfaces using a common attribute design approach for feature-mapping of a state. Specifically, we were interested to see whether participant performance in interpreting the particular scene parameter would be similar for both interfaces, implying potentially strong *design claims*, or guidelines, that could useful for other notification systems. We present results for three potential claims for use of: background shading to convey time since state changes, metaphoric state representations, and selective use of color.

Both Smiling George and the Spinning Cube conveyed the amount of time since a state change by slowly darkening the background, while the Activity Graph simply used one of the six bars. A scene recall question testing participant understanding of this attribute showed this technique to be less effective than the progressively increasing time bar on the Activity Chart.

There were at least three instances of strong metaphors used similarly in the Smiling George and the Spinning Cube interfaces, each conveying: movement activity, position within the room, and numbers of scene. Movement activity was expressed metaphorically—each used movement of the object of interest (lateral to circular and rotation speed) to convey the amount of physical activity within scene. Based on scene recall performance, the lateral and circular motion should to be much more effective than the simple bar chart, although rotation was interpreted poorly. Likewise, the position of actors (portion standing or sitting) was depicted by the height of George's gaze (as if he was looking at the scene actors), the increasing height of the location bar in the graph, and the vertical position of the cube within the display. The question that tested this metaphor supported stronger scene recall than demonstrated on most other questions. Finally, the population level of the scene was represented by the size of the cube (growing as population increased) and George's happiness (degree of smile)—again, the simple activity bar surpassed this metaphor in conveying scene characteristics.

As a final potential design claim, we were interested to see how selective use of color for highlighting specific states would be understood and later recalled. This included two cases: the only use of color change within the cube interface that represented collaboration levels and the red border that was rendered around both the Smiling George and Spinning Cube displays. The sole use of color change within the cube, however, certainly was not effective. While the almost all participants recalled the intruder presence conveyed by the red border, similar levels of high recall were exhibited by participants that were using the graph interface in that scenario, clouding the certainty of this claim.

5.3 Preference Data

For the third hypothesis, we looked at our preference data which consisted of the end of round upside claims that users agreed or disagreed with, as well as the final ability awards, where users picked the best interfaces for a series of criteria.

Aggregating all of the upside claims revealed the average scores below (see Figure 3). An ANOVA test revealed a significant difference among the interfaces with primary tasks ($F(2,1086)=7.68$, $MS=2.08$, $p<.01$) , which was further investigated with t-tests. These found a significant difference between the graph and the cube ($p<.02$) and between the graph and the face ($p<.01$). Among the interfaces without the primary tasks we also discovered differences ($F(2,1119)=27.9$, $MS=2.04$, $p<.01$). T-tests showed significant differences between the graph and cube ($p<.01$) and the graph and face ($p<.01$). Also significant were the differences between the primary task and non-primary task versions of the graph ($p<.01$) and the cube ($p<.05$)..There was no significance for the face's differences. All three interfaces scored higher when the primary task was removed.

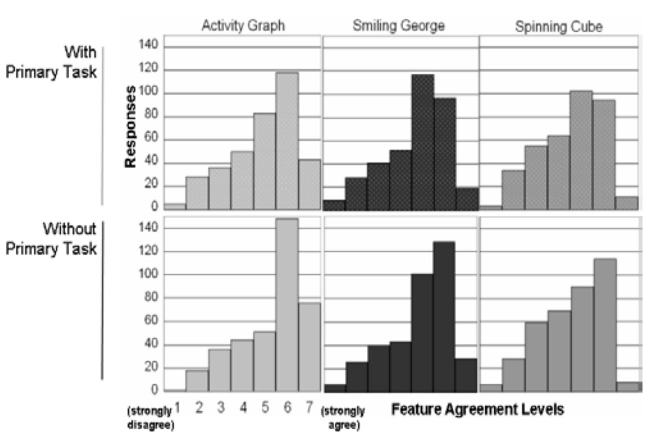

Figure 2: Numbers of participant responses to key interface claims (e.g., *the interfaces provided an overall sense of the information*), assessed after each interface was used for about seven minutes; response numbers for all questions are combined and categorized by agreement level (strongly disagree to strongly agree)

	With primary task		Without primary task	
	Mean	Std Dev	Mean	Std Dev
Graph	4.94	1.50	5.34	1.44
Cube	4.67	1.44	4.89	1.45
George	4.53	1.39	4.56	1.40

Figure 3: User perception of interface features, observed during experience with or without a primary task and reported based on a 7-point scale (strongly disagree=1, strongly agree=7)

The ability awards were consistently awarded to the graph (Figure 4), which received 58% of the total votes, while the cube received 26%, with 16% left for the face. An ANOVA test revealed this difference to be significant ($F(2,195)$, $MS=4.74$, $p<.01$), and t-tests confirmed a significant difference between all pairings of groups, with $p<.01$ for each test. There was no significant difference for any interface between the version of it with the primary task and the one without.

6. DISCUSSION AND CONCLUSION

We were surprised to find no correlation between the user preference data and the performance data. Overall, no version of any interface had a significant impact on the performance of users, yet they still clearly indicated a preference for the graph (much more unanimously than our third hypothesis predicted). It is curious that while users did not perform any better with the graph, they would choose it as the best interface. A possible explanation is user satisfaction of notification systems does not depend on effectiveness or other quantifiable aspects of usability, but instead upon more complex aspects like aesthetics and emotion.

Consistent high scores for grading accuracy and grading rate implied that none of the interfaces were interruptive to the primary task. Low scores on the scene recall questions indicated poor comprehension of information for all three interfaces. Additionally, the presence or absence of a primary task had no effect on the questions, meaning observation of the displays with or without full attention had no effect on the comprehension. We expected that the addition of a distracting and involving task would surely cause more information in the secondary display to

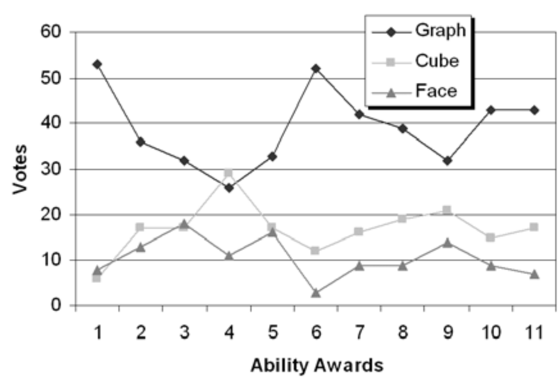

Figure 4: Number of votes for each of 11 ability awards: 1-6 concerning best for mapping each state (as introduced in Sec. 2), 7=easiest to learn, 8=easiest to use, 9=least distracting, 10= easiest to recall, 11=overall satisfaction of use)

be missed, but this was not the case. Consistent poor performance on the scene recall questions made it difficult to extract specific performance claims about the interfaces, because there was not enough contrast among the different questions' scores. High variance found throughout various observations meant that we had a high noise element in our experiment, in part possibly due to a high number of independent variables embedded in the fact that we had fully implemented interfaces with many features. This made it difficult to draw any significant conclusions in this performance data, providing no support for either the first or second hypothesis.

7. FUTURE WORK

In considering the next steps we would take to assess usability of our interface prototypes and other notification systems that support scene monitoring, we recognize two broad possible courses of action:

- **Value user preference indications without placing** immediate concern on performance metrics, focusing design refinement efforts on the Activity Graph and revising future test methodologies that extract more details user preference than task-based performance.
- **Improve the test platform and analysis techniques** (especially activity script files and scene recall questions) to be certain that near-perfect comprehension of scene information can be achieved with unencoded notification delivery (perhaps a simple ticker) in test cases without a primary task; retesting iteratively refined prototypes once verified.

Adoption of either general strategy has important implications for the larger notification systems usability engineering community. The first implies that perhaps our testing objective, at least for formative studies, should not be focused on obtaining performance metrics. Considering the cost associated with preparing the program scripts and software logging necessary for large scale, lab-based performance testing, as well as the relative complexity of the data analysis, testing for preference data implies a considerably less involved user-testing process. If we were to focus on collecting preference data, we would employ a participatory design technique, perhaps encouraging users to think aloud or even use a more fully developed system in their natural work environments. If we insist on the importance of task-based performance data, we must be certain about the validity of our test platform. Development and validation costs for such a platform are quite high, especially for typical usability budgets. However, the research community can support this practical requirement by developing and recognizing general testing protocols for typical application design goals (such as the broad class of displays conveying scene monitoring information). Proven, generic test platforms should be readily available for use, providing low cost, easily comparable, indisputable inference about specific designs.

Of course, our experience described here is based on only a single observation. It would be interesting to collect similar cases—perhaps most of which are not reported in literature—to determine the scope of this dilemma and set a course for the future.

8. REFERENCES

[1] Bartram, L., Ware, C., & Calvert, T. (2001), Moving icons: Detection and distraction, *in Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT '01)*, IOS Press, pp.157-165.

[2] Bowden, R., Kaewtrakulpong, P. & Lewin, M. (2002), Jeremiah: The face of computer vision, *in Proceedings of the 2nd International Symposium on Smart Graphics*, ACM Press, pp. 124-128.

[3] Boyle, M., Edwards, C. & Greenberg, S. (2000), The effects of filtered video on awareness and privacy, *in Proceedings of the Conference on Computer Supported Cooperative Work (CSCW '00)*, ACM Press, pp. 1-10.

[4] Cadiz, J.J., Venolia, G., Jancke, G. & Gupta, A. (2002), Designing and deploying an information awareness interface, *in Proceedings of the Conference on Computer Supported Cooperative Work (CSCW '02)*, ACM Press, pp. 314-323.

[5] Catrambone, R. & Stasko, J., & Xiao, J. (2002), Anthropomorphic agents as a user interaface paradigm: Experimental findings and a framework for research, *in* W.D.Gray & C.D. Schumm (eds.), *Proceedings of the Twenty-fourth Annual Conference of the Cognitive Science Society (CogSci '02)*, Erlbaum, pp. 166-171.

[6] Chewar, C.M., McCrickard, D.S., Ndiwalana, A., North, C., Pryor, C., & Tessendorf, D. (2002), Secondary task display attributes: Optimizing visualizations for cognitive task suitability and interference avoidance, *in Proceedings of the Symposium on Data Visualisation (VisSym '02)*, Eurographics Association, pp. 165-171.

[7] Cutrell, E., Czerwinski, M., Horvitz, E. (2001), Notification, disruption, and memory: Effects of messaging interruptions on memory and performance, *in Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT '01)*, IOS Press,, IOS Press, pp.263-269.

[8] Dourish, P. & Bly, S. (1992), Portholes: Supporting awareness in a distributed work group, *in Conference Proceedings on Human Factors in Computing Systems (CHI '92)*, ACM Press,541-547.

[9] Erickson, T. & Kellogg, W. (2000), Social translucence: An approach to designing systems that support social processes, *ACM Transactions on Computer-Human Interaction (TOCHI) 7*(1), 59-83.

[10] Greenberg, S. & Kuzuoka, H. (2000), Using digital but physical surrogates to mediate awareness, communication and privacy in media spaces, *Personal Technologies 4*(1), 1-17.

[11] Hudson, S.E. & Smith, I. (1996), Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems, *in Proceedings of the Conference on Computer Supported Cooperative Work (CSCW '96)*, ACM Press, pp. 248-257.

[12] McCrickard, D.S., Catrambone, R., Stasko, J. T. (2001), 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 '01)*, IOS Press, pp.558-565.

[13] McCrickard, D.S & Chewar, C.M. (2003), Attuning notification design to user goals and attention costs. To appear in *Communications of the ACM*, March 2003.

[14] Pedersen, E.R. & Sokoler, T. (1997), AROMA: Abstract representation of presence supporting mutual awareness, *in Conference Proceedings on Human Factors in Computing Systems (CHI '97)*, ACM Press, pp. 51-58.

[15] Redström, J., Skog, T. & Hallnäs, L. (2000), Informative art: Using amplified artworks as information displays, in *Proceedings of DARE 2000 on Designing Augmented Reality Environments*, ACM Press, pp. 103-114.