Designing Visual Notification Cues for Mobile Devices

Christopher S. Campbell
IBM Almaden Research Center
650 Harry Road, 4CP/E3-232
San Jose, CA 95120-6099
ccampbel@almaden.ibm.com

Peter Tarasewich
College of Computer and Information Science
Northeastern University, 161CN
Boston, MA 02115
tarase@ccs.neu.edu

Abstract

Mobile and wearable devices place enormous constraints on input and display form factors as well as on user attention. The key to designing micro-displays is knowing what sizes and configurations are viable for keeping users informed, what flexibility different micro-displays provide for different types of messages, and the learning requirements on the user. An experiment was performed to measure user learning and comprehension of increasing amounts of information on a simulated three-light visual display. Users were required to learn five sets of messages of increasing information and complexity using the small display. Results show that micro-displays can transmit detailed, information-rich messages up to 6.75 bits with minimal training (i.e., few trials and short time frames).

Categories & Subject Descriptors: H5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Interaction styles*

General Terms: Experimentation; Design

Keywords: Handheld Devices and Mobile Computing, Ubiquitous Computing, Notification Cues, Visual Displays, Learning, Comprehension, User Interface Design

INTRODUCTION

Information management is becoming increasingly difficult and complex in mobile environments. People must juggle a multitude of dynamic sights, sounds, and other stimuli that convey information and compete for their limited attention. One way to reduce information overload is through the use of meta-information, which can require less effort to process and can result in fewer or less severe disruptions. If meta-information is deemed important, the person receiving it can make a decision whether or not to seek additional details. For example, mobile workers may not need or want the entire contents of a message every time one becomes available. It may be too distracting (and perhaps too dangerous) to the workers' primary tasks. However, they may wish to receive a notification that a message is available, along with an indication of its importance and source. That way, workers can make their own decision, based on their current situation, whether or not to stop their primary task to access the contents of the message.

This research investigates the design and use of visual notification cues, which indicate the status or availability of information that is of interest to a particular user. Specifically, this paper presents the results of an experiment that measured the learning and comprehension of a visual notification display that conveyed increasing amounts of information to the user. Three lights were used, each with three colors and two levels of intensity. A previous study [6] showed that a three-light design (compared to four other designs with more or fewer lights) was a good choice for conveying notifications on small devices. But that study tested only a fixed amount of information. This experiment extends that work and investigates whether or not larger amounts of information can be conveyed using the same three-light design, and how well people can learn to use the notification cue itself.

BACKGROUND

With mobile applications, there can be a significant number of people, objects, and activities vying for a user's attention aside from the application or device itself [5]. An environment that consists of too many distractions can be confusing and unmanageable. Notification cues have to be designed and used such that they minimize the possibility of overloading the attention of the intended recipient and any surrounding people. Otherwise, the cues may prove to be ineffective or may be ignored completely.

Much effort [3, 7] has been devoted recently to studying notification systems in the form of secondary (or peripheral) displays, which provide information that is not central to a user's primary task. For example, news headlines may scroll across a one-line display on a larger screen. Other research has investigated notification systems and devices specifically for mobile environments. Wisneski [8] described a subtle and private notification device in the form of a watch that changes temperature as stock prices change. Holmquist, Falk, and Wigström [2] tested a device called the "hummingbird" that notified its user of the close proximity of other group members by producing a sound ("humming") and listing identities of the group members. Hansson and Ljungstrand [1] created a "reminder bracelet", worn on the user's wrist, which notified a user of upcoming events. The bracelet consisted of three red LEDs that were triggered progressively as an event drew closer.

CHI 2004, April 24-29, 2004, Vienna, Austria.
ACM 1-58113-703-6/04/0004.

EVALUATING INCREASING INFORMATION AMOUNTS

A good mobile notification design should quickly and completely inform users on a small form factor without requiring a lot of attention or training. While small screens exist (e.g., for watches), lower information rate displays such as LEDs have the benefit of (a) requiring less cognitive effort to understand (i.e., less distraction), (b) allowing micro form factors (e.g., jewelry), and (c) using less power. The goals of the present study were to determine (1) how well users can progressively learn increasingly complex messages on a three-light display, and (2) how much information can be conveyed successfully and consistently on that display.

Information Mapping Functions

Information was mapped to the same display---i.e., three circular pixels or lights in a row---in all test cases. Each light could show red, blue, and green at one of two intensity levels (dim or bright). This means that we could theoretically encode six pieces of information on a single light (3 colors x 2 intensities). With three lights, this display can encode a maximum of 216 (6x6x6) different messages. For this experiment, we chose five message sets, each with more messages (more information) than the last. The messages, based on one or more categories shown in Table 1, were mapped into the cue display using position, color, and intensity.

Table 1. Cue categories and associated values. Combining all values in all possible ways results in 108 different messages.

Category	Possible Values
Source	family, friends, work
Medium	email, voicemail
Type	new, reply, forwarded
Length	long, short
Priority level	high, medium, low

Mapping 1

Here, all three lights were lit with the same high-intensity color. Color represented the source of the message; red for family, blue for friends, and green for work. This mapping used the lights to represent a total of three messages.

Mapping 2

The three lights were the same used in Mapping 1. This time, however, color intensity also varied. High intensity for a given color indicated an email message. Low intensity indicated voicemail. For example, high-intensity blue lights indicated an email message from friends. This mapping used the lights to represent a total of six (3x2) messages.

Mapping 3

The three lights were used, but each with three high-intensity colors (red, blue, green). The left light indicated source, the center light indicated type (new, reply, forwarded), and the rightmost indicated priority (high, medium, low). Each light was lit for each notification. For example, "blue green red" indicated a forwarded message from friends with high priority. This mapping used the three lights to represent twenty-seven (3x3x3) messages.

Mapping 4

The lights were used as in Mapping 3. In addition, two intensity levels were used with the left light (source) to indicate medium (email, voicemail). For example, "blue (low intensity) green red" indicated a forwarded voicemail from friends with high priority. This mapping used the three lights to represent fifty-four (6x3x3) messages.

Mapping 5

Mappings were the same as in Mapping 4, with the addition of two intensity levels for the center light (type) to indicate length (long, short). For example, "blue (low intensity) green (high intensity) red" indicated a long forwarded voicemail from friends with high priority. This mapping used the three lights to represent 108 (6x6x3) messages.

These five mappings were used to create five message-sets. According to information theory, the amount of information in a message is related to the number of possible alternative messages – i.e. the more alternatives, the greater the information. Information is measured in bits – the number of binary decisions needed to identify a single message out of all possible alternatives. This is represented by:

$$H = \log_2 N \qquad (1)$$

where N is the number of alternative messages in the message-set [4]. Information loads range from 1.58 to 6.75 bits for message-sets 1 through 5.

METHODOLOGY

Fifty-two undergraduate and graduate students (forty-four male and eight female) participated in this study. Ages ranged from eighteen to thirty years with an average age of twenty-three. None reported themselves as colorblind. The experiment was conducted on Pentium-4 computers running Windows XP with screen resolutions of 1024 x 768. A Java program presented the cue displays as GIF files of 555 x 250 pixels. Status indicators showed the elapsed time for that trial and the correct answers given for that session.

Design

The design was a one-factor (message-set) repeated measures design with five levels. Dependent measures included number of trials to criterion, time to criterion, response time per trial, and first-click response time. The number of trials per block varied according to how many trials it took to reach the criterion performance level (90%). Messages were selected for presentation in random order without replacement within each block.

Procedure

Each subject completed five task sessions of increasing complexity that involved identifying the message given by the three-light display. Subjects were presented with message-set 1 in the first session and progressed through to message-set 5. At the beginning of each session, a subject was shown a visual explanation of how information from a cue mapped to a specific visual display. When the subject was ready to proceed, the three lights were displayed. Subjects responded by selecting one or more buttons on the screen corresponding to the message represented by the lights. Subjects then received feedback about the correctness of the response, and moved onto another cue in that session when ready. Subjects had a maximum of eight seconds to respond to each cue; otherwise, the cue timed out and was counted as incorrect. Subjects continued with a particular session until they got 90% of their responses correct, at which time they proceeded to the next session. Subjects that completed all five sessions were given US$5 as payment for their performance (otherwise, there was no payment for participation).

During each session, buttons with the answer choices were listed in columns at the bottom of the screen. Each column contained the possible values for a category from Table 1. Only those categories relevant to a mapping were listed. Once a selection from each column was made, or when the question timed out, the program highlighted the correct answer on the buttons. The percentage of correct answers for each session was displayed after each response, but the first determination of whether or not to proceed to the next session was made after the first ten answers. Thereafter, the percentage correct was calculated after every answer on a moving basis over the ten most recent responses.

RESULTS

Results were analyzed by calculating the number of trials and time to reach criterion for each set of messages. The number of trials was simply a count of trials in each condition that were performed before the running average reached 90% correct or greater. Because the running average was calculated over a window of ten trials, the lowest number of trials in a condition is ten. Time to reach criterion was calculated by summing the times of all trials in a condition. As the message-set factor was within-subjects, all ANOVAs were performed using a repeated measures analysis and t-tests were performed using paired samples.

A one-way repeated measures ANOVA showed a reliable increase in the number of trials needed to reach criterion across conditions (F(4,180) = 8.30, p < 0.001). Figure 1 shows the mean number of trials to criterion plotted against the information transmitted (in bits) for each message-set. We see that performance remains at ceiling until around 5.0 bits and then falls rapidly (i.e., the number of trials to reach criterion increases rapidly).

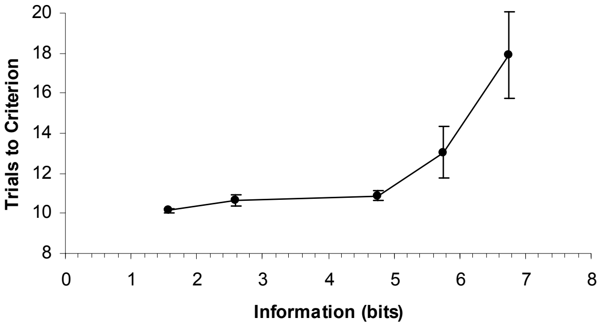

Figure 1. Mean number of trials to reach criterion performance across information levels (±SE).

A one-way repeated measures ANOVA showed a reliable effect of message-set on the time to criterion (F(4,180) = 26.04, p <0.001). Similar to Figure 1, time to criterion increases steadily across message-sets from a mean of 3 minutes to almost 19 minutes.

Looking at the results broken-down by trial, Figure 2 shows the learning curves over the first 10 trials by message-set. Little learning occurs for message-set 1 because initial performance is already above criterion. For message-set 5, however, few subjects have reached criterion by trial 10.

Total response time averaged over trials reliably increases across message-sets (F(4,180) = 502.85, p < 0.001). One could argue that the increase in response time for increasing message-sets is solely the result of having to press more buttons. Therefore, the total first-click response time to criterion was analyzed. This is calculated as the time from stimulus presentation until the first button is pressed. A one way repeated measure ANOVA indicated a reliable increase in first-click time across message-sets (F(4,180) = 7.93, p < 0.001). Subjects took about 17 seconds total before giving a response for message-set 1, but took up to 36 seconds for message-set 5.

Analysis of time-outs – the trials that timed-out due to no response before the time limit – showed that very few time-outs occurred during the experiment. Only 9 trials total timed-out and there was no reliable difference across message-sets (F(4,180) = 1.71, NS).

DISCUSSION

The results show very good performance by many subjects as performance was near ceiling for message-sets 1, 2, and 3. In other words, subjects had no trouble learning up to 4.75 bits of information in 10 trials or less. To learn all 6.75 bits of information or 108 alternatives required only 19 trials on average. Performance, however, began to decline more dramatically after about 5 bits of information. Also, response times showed that this high accuracy was achieved at the expense of time. Essentially, the response times increase steadily over message-sets and significantly increase across message-sets 2 and 3 (from 35 seconds to 50). This shows that there is a cognitive cost to learning

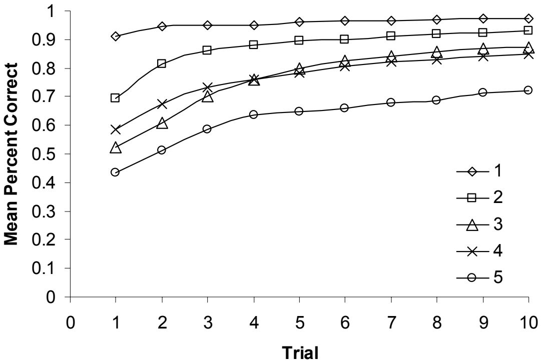

Figure 2. Mean percent correct across trials for message-sets 1 – 5. SE bars omitted for clarity.

larger amounts of information from the same size display even if this cost is not reflected in the accuracy alone.

The argument could be made that these effects are due to the increased number of buttons that need to be clicked across message-sets. However, looking at first-click response times – the time from display presentation to the first click – there is a significant increase in response time ($F(4,180) = 7.93$, $p < 0.001$). Because this time is summed across trials, the increase could be the result of increased number of trials and not increased times on each trial. However, an analysis in which first-click times were averaged over trials still showed a reliable increase in first-click response times as well ($F(4,180) = 17.16$, $p < 0.001$). Thus, having to click additional buttons is unlikely to be the cause.

CONCLUSIONS AND FUTURE WORK

The guiding question of this study was how much information can be displayed by low-information-rate micro-sized displays? Are designers stuck with one or two bit messages or can a greater range of messages be conveyed? The results of this study indicate that people can quickly learn fairly large notifications of over six bits with only three pixels – even with a response time limit of eight seconds. This makes low-information-rate, micro displays practical for consumers who are not willing to endure a long training cycle. Design possibilities are enhanced as well because many message schemes could be employed with over six bits of information.

Clearly, there are a number of factors that contributed to finding robust performance over increasing information rates. Among them is the organization of the message-sets into categories or chunks. For example, the 108 messages of message-set 5 could be decomposed into five categories. So, instead of having to identify one out of 108 unrelated messages, the subject only needs to identify 5 categories with 2-3 alternatives per category. Future work will be aimed at comparing message-sets that can be organized by category against those that can not and also against those in which the user provides a customized organization. This comparison will allow designers to determine the limitations of low-information rate displays for less structured information.

Designing message-sets that are hierarchical may provide a particular advantage---i.e., allows people to progressively learn instead of trying to memorize the large 108 message set at one time. This appears to increase the size of messages that can be conveyed and reduce the learning requirements. It also has the benefits of (a) allowing for immediate use of the notification system with almost no learning and (b) providing advanced functionality for expert users who wish to maximize the utility of the display.

Future work will also compare several possible methods for learning a set of notifications. It would be valuable to know, for example, if providing hierarchical information-cue mappings facilitates learning when presented in (a) progressively more difficult training blocks, (b) random blocks, or (c) all in one difficult block. Another valuable comparison is design versus customization. In other words, if people are allowed to design their own information-cue mappings, will the notification system be easier to use? Customization also addresses privacy and security concerns. For example, three red lights on a ring, even when noticed by other people nearby, could convey a message only understood by the wearer.

ACKNOWLEDGEMENTS

The authors would like to thank Jun Gong and Payal Kapoor for their help with administering this experiment.

REFERENCES

1. Hansson, R. and Ljungstrand, P. The Reminder Bracelet: Subtle Notification Cues for Mobile Devices. *Ext. Abstracts CHI 2000*, ACM Press (2000), 323-324.
2. Holmquist, L. E., Falk, J. and Wigström, J. Supporting Group Collaboration with Inter-Personal Awareness Devices. *Personal Technologies*, 3(1-2), 13-21, 1999.
3. Maglio, P. P. and Campbell, C. S. Tradeoffs in Displaying Peripheral Information. *Proc. CHI 2000*, ACM Press (2000), 241-248.
4. Shannon, C. E. & Weaver, W. *The Mathematical Theory of Communication*. University of Illinois Press (1949).
5. Tarasewich, P. Designing Mobile Commerce Applications. *Comm. of the ACM*, 46(12), 57-60, 2003.
6. Tarasewich, P., Campbell, C., Xia, T., and Dideles, M. Evaluation of Visual Notification Cues for Ubiquitous Computing. *Proc. UbiComp 2003*, Springer-Verlag (2003), 349-366.
7. Van Dantzich, M., Robbins, D., Horvitz, E., Czerwinski, M. Scope: Providing Awareness of Multiple Notifications at a Glance. *Proc. AVI 2002*.
8. Wisneski, C. A. *The Design of Personal Ambient Displays*. M.Sc. Thesis, May 1999, MIT Media Laboratory, Cambridge, MA.