Evaluation of Visual Notification Cues for Ubiquitous Computing

Peter Tarasewich[1], Christopher S. Campbell[2], Tian Xia[1], and Myra Dideles[1]

[1] College of Computer and Information Science
Northeastern University
360 Huntington Avenue, 161CN, Boston, MA 02115
{tarase,tianxia,cdideles}@ccs.neu.edu
[2] IBM Almaden Research Center
650 Harry Road, NWE/B2-229, San Jose, CA 95120-6099
ccampbel@almaden.ibm.com

Abstract. With increased use of mobile information technology and increased amounts of information comes the need to simplify information presentation. This research considers whether low-information-rate displays (such as those used in mobile devices) can provide effective information awareness. An experiment was performed to measure the performance/size tradeoff of visual displays ranging in size from two LEDs to nine LEDs, and using a number of display characteristics – i.e., color and blinking in various combinations. Results show a reliable tradeoff between performance (participant response time and accuracy) and display size (number of LEDs). However, even the full set of 27 messages can be conveyed with high recognition accuracy using only three LEDs by mapping the messages into color and position. Thus, mobile devices with micro-level form factors can be designed to convey critical information and provide effective notifications. Future work and a prototype developed from this work are discussed.

1 Introduction

With the growing use of mobile information technology comes an increased concern about the impact of such technology on society as a whole. Many computer-based devices are no longer constrained to relatively permanent work or home environments, but are likely to be found in almost any physical location or in any social setting. Concerns with mobile applications have arisen in areas such as safety (e.g., using a handheld organizer while driving) and appropriateness (e.g., using a cell phone in a restaurant). Many organizations have even started regulating the use of mobile devices in specific locations and under certain circumstances, sometimes using technology itself (such as cell phone jammers) to enforce “proper behavior”. With an increase in mobile devices also comes an increased risk of “attention overload,” which can occur when individuals are overwhelmed and interrupted by intrusive and attention demanding external events, such as loud auditory cues or flashing bright lights [4].

A.K. Dey et al. (Eds.): UbiComp 2003, LNCS 2864, pp. 349–366, 2003.

In such technology-laden environments, information overload is a serious problem as well. While information is necessary to perform many tasks, the human mind is limited in terms of how much information it can process at one time. The problem of information management becomes even more 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, a mobile worker may not need or want the entire contents of a message or an announcement every time one becomes available. It may be too distracting (and perhaps too dangerous) to the worker's primary tasks. However, they may wish to receive a notification that a message is available, along with an indication of how important it is, and its source. That way, the worker 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 notification cues, which indicate the status or availability of information that is of interest to a particular user in a ubiquitous setting. Notification systems must "present potentially disruptive information in an efficient and effective manner to enable appropriate reaction and comprehension" [16]. Examples of notification cues include the ringing of a cell phone for an incoming call and the chime on a handheld device when a text message arrives. Notification cues are a form of information, and questions arise such as what form these cues should take and how appropriate they are in different settings. Determining notification cues for use in ubiquitous environments can become quite complex, requiring the selection of appropriate delivery channels based on continuously changing contexts and dynamic information needs.

Specifically, this paper presents the results of an experiment that measured the comprehension of and preferences for different visual displays. Each display conveyed the same amount of information, but differed in the number of lights (LEDs) used, their physical arrangement (pattern), the colors used, and whether the lights blinked or not. More lights used in a display means that more room is needed on a device that conveys the notification cue. More lights, however, and the pattern of those lights, may convey information more quickly and easily to the user. Blinking lights and different colors may add complexity to the cue, and may make it more difficult to understand the cue, but allow for a more efficient (i.e., compact) display. This experiment, therefore, investigated whether or not tradeoffs existed between several different cue displays.

The paper is organized as follows. Section 2.0 provides some background on notification cues, including material on human attention and distractions. Section 3.0 describes the methodology used in the experiment, including a description of the configurations and reasons behind the choices made for each display. Section 4.0 presents the results of the experiment, and section 5.0 discusses the results. Section 6.0 draws some conclusions and provides directions for future research.

2 Background

Notification cues can take on various characteristics. Cues can be visual, auditory, tactile, or multimodal in nature. They can be private such that only the receiver is aware of then, or public such that everyone in the immediate vicinity will receive the cue. Cues can also range from being quiet and subtle to being loud and intrusive. A ringing cell phone is an auditory, intrusive, and very public notification cue. A vibrating cell phone is a tactile, subtle, and private notification cue that can convey the same information [4]. There may be different situations where the use of each of these cues is more appropriate. Notification cues also need to safely compete for a user's attention in a world full of an increasing number of distractions, especially in the mobile environment. Important information needs to reach its intended user quickly. Critical notifications that are delayed too long or fail to be delivered can lose their value completely.

The design and use of notification cues must take into account the intricacies of human attention in dynamic environments. Attention involves the allocation of perceptual or cognitive resources to something at the expense of not allocating them to something else [5]. Humans have a limited amount of resources available for allocation to different tasks, therefore everything cannot be attended to at once. People can attend to a modality (vision, hearing, touch, taste, smell), a color, a shape, or a location [5]. The decision to attend specifically to one of these over the others arises from the task at hand. However, events occurring in the unattended modalities will not go unnoticed [15, 18]. For example, when reading a book, a person may ignore most sounds but will respond if their name is called. The brain processes unattended modalities at a level that allows for recognition but keeps them unnoticed for the most part [23].

With computer applications that are used in the office, home, or similar settings, the context is known and is relatively stable from minute to minute. While this does not mean that there cannot be multiple activities competing for a user's attention (e.g., animated ads, intrusive pop-under windows, stock alerts, and email notifications), the environment outside of the computer is fairly consistent for a given user from day to day. Most offices and homes function with a fair amount of regularity and predictability, even if they do experience a great amount of activity. The user can devote a relatively consistent amount of attention to actually performing tasks on the computer.

On the other hand, with mobile applications, there can be a significant number of additional people, objects, and activities vying for a user's attention aside from the application or computer itself [21]. Furthermore, since devices are completely mobile, this outside environment can change rapidly from moment to moment. A mobile application may not be the focal point of the user's current activities [7], as the user may be trying to juggle interaction with a mobile device along with other elements in the environment (e.g., walking on the sidewalk of a busy city street with small children while receiving directions from a navigation system). Mobile activities can be complex because of changing interactions between the user and the environment. The amount of attention that a user can give to a mobile application will vary over time, and a user's priorities can also change unpredictably [11].

An environment that consists of too many distractions can be confusing and unmanageable. Hansson, Ljungstrand, and Redstöm [4] defined *attention overload* as a situation where "people are overwhelmed and interrupted by obtrusive and attention demanding external events, such as loud auditory cues or flashing bright lights." 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 research effort has been devoted recently to studying notification systems in the form of secondary displays, or peripheral displays, which provide information to the user that is not central or critical to their current or primary task. For example, a user may have a one line display on their computer screen in which scroll current news headlines. Studies have looked at the effectiveness of presenting information in secondary displays in various formats [1, 10, 14, 16, 24]. The Scope notification summarizer [24] used color and relative location to convey the source, priority, and type of information available to the user. Research has generally found that user performance on primary tasks is negatively impacted by these secondary tasks [1, 14], with some exceptions [16].

Other research has investigated notification systems and devices specifically for mobile environments. Wisneski [27] described a subtle and private notification device in the form of a watch that changes temperature as stock prices change. Holmquist, Falk, and Wigström [8] 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 the identities of the group members.

In a first attempt at creating a subtle notification cue that was also public, Hansson and Ljungstrand [3] created a "reminder bracelet", worn on the user's wrist, which notifies a user of upcoming events (e.g., meetings). The bracelet consists of three light emitting diodes (LEDs) that are triggered progressively as an event draws closer. The notification information comes from a PDA (carried by the user) that is wired to the bracelet. With this device, other people interacting with the user can clearly see that the user is being notified about something.

The delivery of notifications can sometimes be improved by interpreting the current context of the user through the use of sensors and by determining the priority of the information being sent. Nomadic Radio [19] is an auditory notification device that manages voice and text messages in a mobile environment. The form of notification is chosen based on the content of the message, whether or not the user is speaking, and the user's responses to previous messages. Horvitz, Jacobs, and Hovel [9] described the concept of attention-sensitive alerting, whereby the costs of a user's interruption are balanced with the costs of deferring alerts. Horvitz, Jacobs, and Hovel looked at probabilistic models that can make inferences about a user's attention under uncertainty. In addition to this, they created a system to automatically classify the criticality of the alert (in this case, email messages).

Schmidt et al. [20] have experimented with a cell phone that changes its ring type (vibrate, quiet ring, loud ring, silent, or mixed-mode) based on the context of the phone. A sensor board was created to measure light level, tilt, vibration, proximity, temperature, pressure, and sound level. The readings from the board were interpreted by software to provide information current situation of the phone, and its ring was adjusted accordingly. For example, if the phone was judged to be on a table, it would ring quietly (the assumption being that the user is in a meeting).

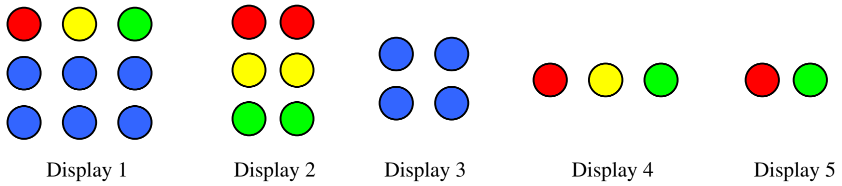


Fig. 1. The set of cue displays (low-information-rate displays) used in the experiment. Display 5 uses multi-color LEDs (red/yellow/green)

3 Evaluation of Low-Information-Rate Displays

One key issue in designing mobile notification displays is how to adequately inform users given a small or even micro form factor without requiring a great deal of attention or needing long training sessions to learn coded messages. While very small LCD screens have been developed (e.g., watch displays), lower information rate displays such as LEDs have the benefit of (a) requiring less cognitive effort to understand (i.e., less distraction), (b) allowing for smaller and even micro level form factors, and (c) using less power. Simply speaking, the less information conveyed, the less attention required to use that information. However, less information does not mean the message is not informative. Even small amounts of critical information can be highly informative and keep the user aware in mobile situations.

An experiment was conducted to test comprehension and preferences of notification cues across five different cue displays (i.e., low-information-rate displays) ranging in number of LEDs from two to nine (see Figure 1). The number of LEDs used is related to the size of the display and the complexity of the display is related to how many different states each LED can assume. As complexity increases, performance should decrease. Because the same number of messages must be conveyed by all the cue displays, complexity will increase as size decreases. This is a critical tradeoff for mobile displays that attempt to minimize size. The goal of the present experiment was to determine the function of this tradeoff and find the optimal point on this function. This optimal point will show the smallest array of LEDs that can be used while still maintaining high performance and maximizing preference. In other words, we assume that as size decreases, complexity increases, and performance will decrease. If the relationship between size and performance is linear, then no optimal point can be found. A non-linear relationship with a knee in the curve, however, would indicate there is an optimal design point that could be taken into account during system design.

Identical messages were mapped into each cue display and consisted of three cue dimensions at three levels each for 27 distinct messages (see Table 3). The messages were mapped into the cue displays using position, color, and blinking.

Subjects were given a pre-experiment questionnaire concerning current mobile technology use and demographic information. Subjects were then presented with each cue display showing one of the twenty-seven messages and asked to indicate the priority, source, and descriptor of the message. Response accuracy, start time, and end time were recorded. Finally, subjects were given a post-experiment questionnaire to indicate which cue display they preferred.

3.1 Complexity and Mapping Functions

Assume a *cue display 0* consisting of five LEDs in a single row. Each LED has ON and OFF states, corresponding to the numbers “1” and “0”. If we think of binary numbers represented by these five LEDs, the following mapping function is possible:

Table 1. Example of a complex mapping as demonstrated by a binary coding of messages across five LEDs in a row

Binary code	Cue value (Priority level, Source, Descriptor)
00001	High, family, reminder
00010	High, family, news
00011	High, family, email
00100	High, friends, reminder
…	…
11011	Low, work, email
00000	Default state: no information is conveyed
11100 – 11111	(Not used)

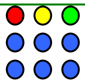
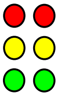
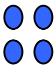
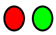
This kind of mapping for conveying meta-information is inappropriate, not only because it is difficult for most people to remember all 27 binary numbers, but also because the mapping between binary numbers and cue values is not intuitive. Interpreting cue display 0 requires scanning through all the binary codes in long-term memory until the correct one is found. This requires scanning one to 27 codes. Before scanning the codes, they need to be recalled from long-term memory and loaded into working memory. Unfortunately, the capacity of working memory is fixed, about five to nine independent items, according to Miller [17]. This means that we could reduce the interpretation time by reducing the information we need to access. This can be achieved through the grouping of related information in cue display designs.

In contrast to the design of cue display 0, which mixes all three categories together and generates 27 instances, we will keep these three categories separate, and interpret the display using three mapping functions corresponding to these three categories. By grouping the information by category, we will have at most nine instances from three functions. We interpret one value for each of three categories, and get a whole cue value simply by combining three values together. Good mapping functions should lead to good display interpretation times.

3.2 Displays Used for the Experiment

Each display conveys the same amount of information, but differs in terms of the number of lights (LEDs) used, their colors, whether they blink or not, the configuration pattern of the lights, and how the information presented on the display is mapped to these characteristics (see Table 2). Using more lights means that more space is taken up by the display that conveys the notification cue. More lights, however, and the layout of those lights, may convey information more quickly and easily to the user. Blinking lights and different colors may add complexity to the cue, and may make it more difficult to understand the cue. This experiment investigates whether or not there are tradeoffs between a variety of visual displays used to present notification cues.

Table 2. Visual qualities and mapping complexity of the five cue displays

Display	No. LEDs	"On"-state Used	"Off"-state Used	Blinking Used	Multi-color LEDs	Mapping Complexity
	9	X				Very Low
	6	X	X			Low
	4	X	X	X		Medium
	3	X	X	X		High
	2	X	X	X	X	Very high

Cue Display 1. The LEDs in the first row are red, yellow, and green; the other LEDs are blue. Values for priority are mapped to the LEDs of the first row, values for source to the second row, and descriptor to the third row. Each position in a row corresponds to a value for that information category. In this example, high is mapped to the first light of the first row, medium to the second, and low to the third. Mappings for source and descriptor are similar. Exactly one LED is lit in each row for any notification cue.

This cue display is the simplest mapping from message to cues. Users do not need to combine several dimensions before comprehending its meaning. The independent items in working memory could be as few as three, because three physically separate elements in the same category that are stored together as a group in long-term memory may be recalled and maintained in working memory as a single entity [26]. It is possible not to store a whole category as one item in memory in this design, because users may not need to process information in each group.

Cue Display 2. LEDs in the first row are red, in the second row are yellow, and in the third are green. The LEDs could also all be the same color. Priority is mapped to the first row, source to the second, and descriptor to the third, as above. This time, however, the row values are indicated by a binary code. Both LEDs lit indicate the first value (e.g., high), just the first LED lit indicates the second, and just the second LED lit indicates the third value.

Cue Display 3. All four LEDs are the same color. The first row is priority, and its value is conveyed through a binary code as in display 2 (11 = high, 10 = medium, and 01 = low). The bottom left LED is source, and the bottom right is descriptor. The values are indicated by blinking, lit, and off (e.g., blinking = family, lit = friends, and off = work). This is different from displays 1 and 2 in that "off" is now used to indicate a value. We could also place all four LEDs in a row.

Cue displays 2 and 3 map the values of each category to the LEDs using more complex encoding functions than the first design. These functions are not straightforward, and the mapping of each basic element needs to be calculated before it is comprehended. So we may not rehearse and maintain 3 basic elements in one category as a single entity. These functions need more working memory, and are assumed to produce a slower response due to the tradeoff between performance and device space consumed.

Cue Display 4. The first position is priority, the second is source, and the third is descriptor. Blinking is the first level of information, lit is the second, and off is the third. While this mapping is fairly intuitive, it requires subjects to associate blinking and on/off together as a single dimension from high to low. Additionally, the blinking may be missed during a hasty response leading to more errors and worse performance.

Cue Display 5. Using color LEDs (e.g., red, yellow, and green in the same LED) one can reduce the number of LEDs down to two. The first LED is message source, with red indicating family, yellow friends, and green work. The second LED is descriptor with the same mapping. Priority is indicated by the blinking of the LEDs (regardless of color). Both LEDs blinking signify high priority, one blinking is medium priority, and none is low. This design does not use grouping of information. We could not use 2 LEDs to represent 3 groups by position alone. Users need to remember all 27 instances of a function. The design will utilize both changing of colors and different blinking rates in order to reach enough complexity to accommodate all combinations.

An attempt was made to use display dimensions in a way that best produces an intuitive mapping. For example, color is an especially effective way of coding nominal sets [13]. Kahneman and Henik [12] and Ware and Beatty [25] pointed out that the human visual system is very effective at distinguishing a small number of distinct colors. Some research [6] points out that red, yellow, and green should be reserved for "Danger", "Caution", and "Safe", respectively. In keeping with this recommendation, red is used for high levels of each cue category (e.g., high priority), yellow for medium, and green for low levels. Finally, only one level of blinking was used as the use of too much blinking, particularly more than two levels, is not recommended [2].

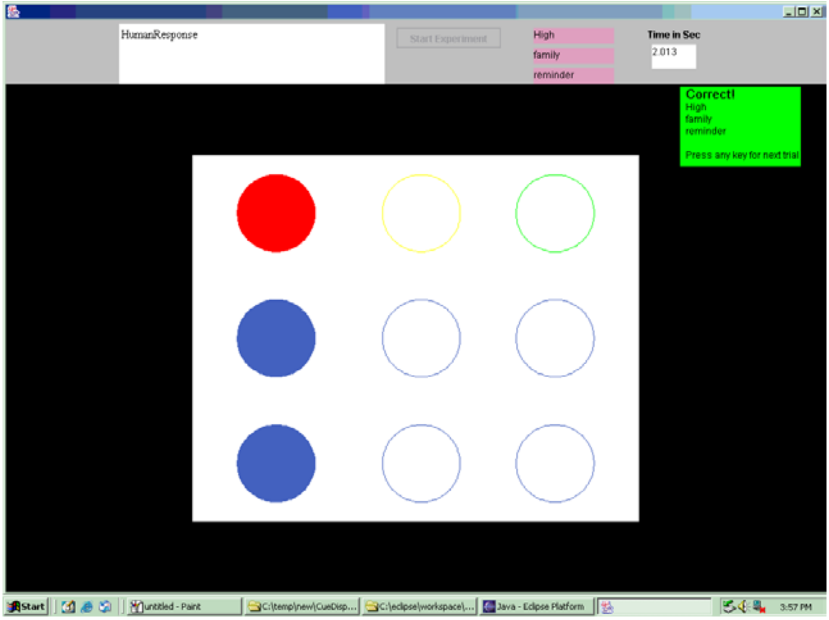


Fig. 2. Screenshot of experiment display, with the stimuli (i.e., cue display) in the center of the screen. The three pink bars at the top of the screen indicate that responses have been made for all three cue categories. Feedback is shown in the green box in the upper-right hand corner indicating that all three responses were correct. Finally, the response time is shown in the white box in the upper right-hand corner

3.3 Method

Subjects. Nineteen college students and one instructor of an information science class at a large university participated in this study. Eighteen were male and two were female. The median age of the participants was 21 years, and each claimed to use a personal computer daily. None of the subjects were colorblind. Nineteen of the twenty subjects carried a cell phone, for 3.2 years on the average.

Design. This was a four-way fully-factorial within-subjects design in which the cue display order was randomized and the trials for each cue display were blocked. In other words, subjects were presented with all the trials of each cue display at the same time (i.e., blocked), but cue display order was random. Cue values were presented randomly without replacement within each block. The four factors included cue display (5 levels), priority (3 levels), source (3 levels), and descriptor (3 levels) with two repetitions for a total of 270 trials per subject.

Equipment and Materials. Subjects completed a pre-experiment questionnaire – a set of questions pertaining to their computer and mobile device background. The experiment was then conducted on a Pentium 4-based personal computer (PC) running Windows XP with the screen resolution set at 1024 x 768 on a CRT screen. A program was written in Java to present the cue displays as animated Graphics Interchange Format (GIF) files. The GIF files were presented on a black background that covered the entire desktop except for the taskbar at the bottom (see Figure 2). The GIF files were 555 X 457 pixels in size, 96 pixels/inch resolution, and had 8-bit color depth. At the top of the screen was a status bar showing what cue responses had been made and the elapsed time for that trial. Subjects completed a post-experiment questionnaire to determine cue display preferences.

Procedure. A training session was first performed where mappings for each of the different displays to be used were shown and explained to each subject. Next, five rounds of testing were performed. In each round, subjects were first shown the mapping for one of the cue displays (randomly chosen) and their responses for the three cue dimensions were recorded. A notification cue conveys meta-information (information about information). For this experiment, the information conveyed by the cues was that summarized in Table 3.

Table 3. Cue categories and associated values. Combining all the values in all possible ways results in 27 different messages

Category	**Possible Values**
Priority level	high, medium, low
Source	family, friends, work
Descriptor	reminder, news, email

Using all three information categories, there are 27 (3x3x3) possible notification cue values that can be conveyed on any display. These were shown two times each, so a particular display showed a random sequence of 54 notification cue values. These cue values were graphical representations of actual notification cues. They consisted of simulated LEDs, which were simply circles drawn in a particular pattern. “Lit” LEDs were simply fully colored circles. “Blinking” LEDs were circles that alternated between empty and filled. Each display value was shown until the subject entered a response signifying the information represented by the notification cue display, but for not more than 6 seconds. Subjects responded by pressing three keys on the numeric keypad of the PC. Each row of the keypad corresponded to a different information category; row one was priority level, row two source, and row three descriptor. Each of the three keys in each row corresponded to a possible value for that category (e.g., 7 = high, 8 = medium, 9 = low). The first key pressed in each row was taken as the response to that display and could not be changed. At the end of each response, the correct answer was shown to the subject. After the first display, the experiment continued with the remaining four displays (for a total of 270 notification cue values tested per subject). The Java program automatically recorded the subjects’ answers and response times.

4 Results

4.1 Performance

Four measures of performance were collected during the experiment – start-time, completion-time (end-time), difference-time, and percent correct. The start-time was the time that the subject entered the first response on the keypad. The completion-time was the time that the last (third) response was entered. The difference-time was the difference between start-time and end-time. Percent correct was the number of cue dimensions correctly specified. Because there were three responses for each trial, the percent correct was measured as a score between 0 and 3 with 0 meaning no correct responses and 3 meaning all correct responses. A separate four-way repeated measures analysis of variance (ANOVA) was applied to each of these measures. Our alpha rate for all ANOVAs was .01 or 1 percent meaning that the p-value for each test must be less than .01 to be considered significant or reliable.

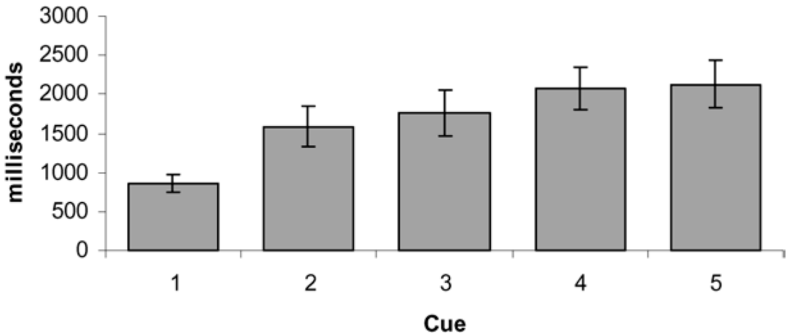


Fig. 3. Start-time increases as mapping complexity and number of LEDs increases

The ANOVA for start-time showed a significant effect of cue display ($\underline{F}(4,72) = 90.23$, $p < .01$) and priority ($\underline{F}(2,36) = 15.66$, $p < .01$). As shown in Fig 3, responses started much faster for cue display 1 (M = 851ms) than the other cues with cue display 5 the slowest. For priority, high (M = 1678ms) and low (M = 1601ms) priorities were reliably faster than medium (M = 1762ms). The error bars in Figures 3 through 6 represent 95% confidence intervals.

The ANOVA for end-time also showed a reliable effect of cue display ($F(4,72) = 72.14$, $p < .01$) and priority ($F(2,36) = 17.68$, $p < .01$) as well as descriptor ($F(2,36) = 7.44$, $p < .01$). Fig 4 shows that cue display 3 has the longest total response time (M = 3645ms) followed by cue display 5 (M = 3568ms) and cue display 4 (M = 3337ms). Similar to start-time, priority for end-time is fastest for high (M = 2982ms) and low (M = 2846ms) than medium (M = 3062ms) priority. Finally, email descriptors (M = 2901ms) were entered significantly faster than reminder (M = 3001ms) or news (M = 2988) descriptors.

The ANOVA for difference-time showed a reliable effect of cue display ($F(4,72) = 34.68$, $p < .01$). In other words, the amount of time it took to enter all three responses varied across cue displays. It took the longest time to enter responses for cue display 3 (M = 1872ms) and the least time to enter responses for cue display 4 (M = 682ms). Also, there was a significant effect of priority ($F(2,36) = 6.74$, $p < .01$) and descriptor ($F(2,36) = 7.24$, $p < .01$).

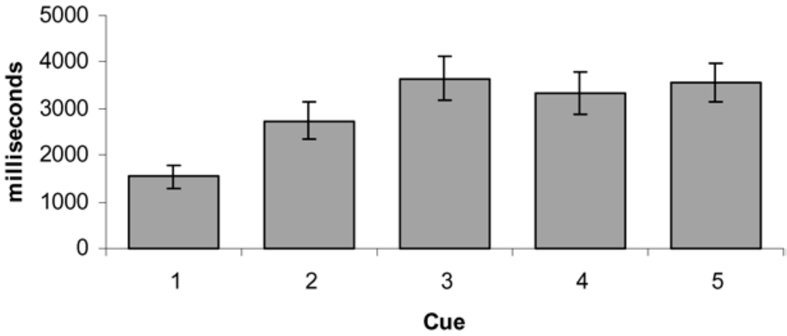


Fig. 4. Completion time is fastest for cue display 1 and slowest for cue display 3

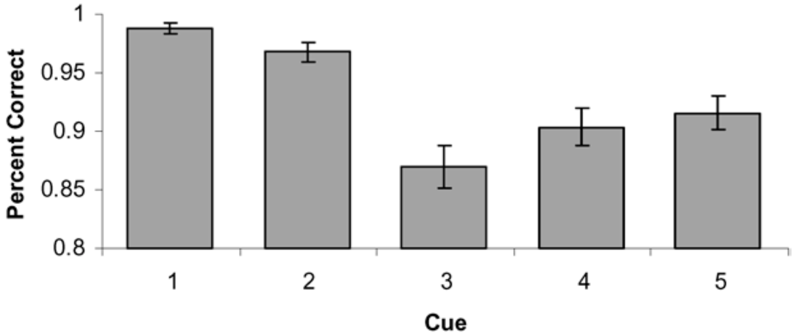


Fig. 5. Percent correct responding is very high for all cue displays but significantly worse for cue display 3

The ANOVA for percent correct showed that only cue display had a reliable effect ($F(4,72) = 6.36$, $p < .01$). Fig 5 shows that cue display 3 had the lowest accuracy while cue displays 1 and 2 had the highest. Overall, the accuracy rates are very high indicating that in the speed-accuracy tradeoff, subjects were focused on high accuracy.

4.2 Questionnaires

Seventeen people ranked the five displays after completing the experimental sessions. In ordering of decreasing preference, subjects preferred displays 1, 2, 4, 3 and 5. Subjects were also asked several more open-ended questions concerning the displays that they had just interacted with. One question asked, if subjects had adequate time to learn any of the displays before using them in the real world, which one(s) would they rather use on a mobile device such as a PDA. Most subjects answered that they would still rather use display one (the nine LED display), with some commenting that the mapping was the most natural and that it was the most easy to use. However, about a third of the subjects commented that displays two, three, or four would be preferable if adequate learning time was available.

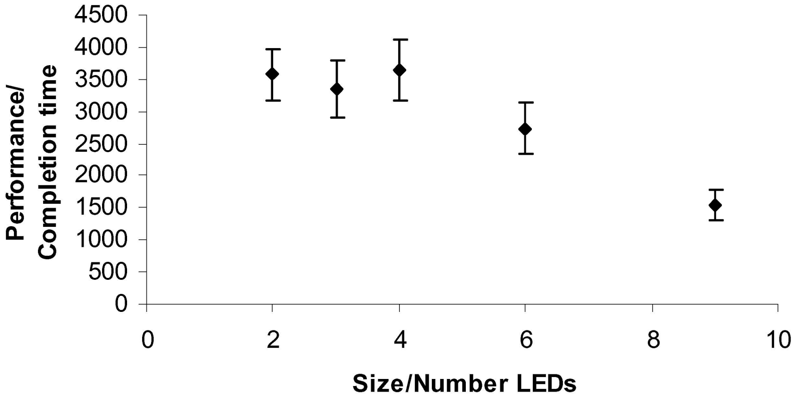


Fig. 6. The performance versus size tradeoff function appears to increase linearly from nine to four LEDs and then plateau

The second question asked that if there was adequate time to learn any of the displays before using them in the real world, which one(s) would rather be used on device such as a watch? For this question, there was a movement in preferences towards smaller displays, with nine people preferring display 4 (three LEDs), three preferring display 5 (two LEDs), and one preferring display 3 (four LEDs). Two still preferred display 1 (nine LEDs), and two display 2 (six LEDs),

The final question asked which displays subjects would prefer not to use again. About half of the subjects commented that they did not like the two LED display (display 5), and a third did not like to use the four LED display (display 3). One person commented that the blinking displays in general were difficult to use.

5 Discussion

As expected, there are explicit tradeoffs between the different displays studied in this experiment in terms of the performance measures collected. What is really surprising, however, is that instead of continuously worse performance, response times flatten out at four LEDs and remain about the same at two LEDs (see Figure 6). In other words, subjects were able to use two LEDs to recognize 27 messages with 92% accuracy, responding in less than 4 seconds – all with little training and only an hour of experience. Part of the explanation for this result is the young subject population used. Certainly, with a wider range of subjects, learning rate could be more of a barrier to using low-information-rate displays.

With cue display 1, subjects performed the best in terms of start time, completion time, and accuracy. This is not surprising, since cue display one was designed to map perfectly with the three message categories and three values for each category. Subjects were able to recognize the cues quickly and accurately. Subjects also seemed to prefer display one overall to the other four displays. Here, the perception may be that if screen real estate is readily available, it should be used for a display that quickly and easily conveys information to the user.

There is a significant increase in both start time and completion time, and a significant decrease in accuracy, for display 2. This display has six LEDs, and while the mapping of its rows corresponds to the information categories, a binary coding system was used to display the three values for each category. Therefore, this increase is not surprising, given the expected extra cognitive processing time that such decoding would incur. Such a display, however, would take up less real estate on a display. A learning curve seemed to factor into the performance with this display, as several subjects mentioned that they would use this one (over display 1) if they had sufficient time to get used to it.

Display 3 was most difficult display for subjects to interact with. Start times were third highest, but completion times were the longest out of any of the displays. Accuracy rates were also the lowest for all of the displays. This display had four LEDs. The top two LEDs were used to represent priority through a binary code, while the bottom two each represented one of the other two categories, with levels represented by on, off, or blinking. There are two possible contributing factors to the unexpectedly high complexity of this display, 1) it combined two different visual representations in one display (binary, blinking) and 2) it used blinking, which seems to be more difficult and time consuming to comprehend by nature. Surprisingly, while many of the subjects did not like display 3, it was not the least preferred display.

While display 4 showed a fairly high start time, its completion time, although higher, was not significantly different than that of display 2. Display 4 consisted of three LEDs, each of which mapped to a category. Levels were indicated by on, off, or blinking. Accuracy rates were not as good as with display 2, but were significantly better than display 3. When subjects were asked which display they would like to use with a watch-sized device (given adequate learning time) most preferred this display.

Display 5 was the display that half of the subjects said they would not want to see again. This two LED display combined the three categories using three colors and binary blinking. The cognitive effort required to process the cues on this display is reflected in the highest start times, and the fourth highest completion times. Accuracy, however, was not significantly different from display 4.

Some of the preferences or performance results found during this study may be related to the characteristics of the display implementations. Priority was always the first row or LED on the displays, so that may explain why the performance was better for priority (or why start time was significantly lower). Priority was also often indicated with color-coding.

Multiple colors are an integral part of all displays except display 3 (the four LED display). The colors may have helped to differentiate the different category values from each other. In the case of display 5, color recognition was necessary to discriminate the different level of two of the categories. When color was used to encode different levels of a particular category, the colors red, yellow, and green were used, similar to a traffic light pattern. These were matched up to priority levels of high, medium, and low, or other category levels matching those going from left to right on the keyboard.

Binary coding was used in displays 2 and 3, where two LEDs were used to convey information about a category level. This may have required additional cognitive effort due to a decoding process needed to match the code to the appropriate level, although an attempt was made to match the codes from left to right with the levels as assigned on the keyboard.

Blinking was also used to indicate category levels in several of the smaller display designs. In displays 3 and 4, LED states of blinking, steady, and off were used to indicate one of the three levels for a particular category. In display 5, blinking of no, one, or both LEDs was used to indicate priority. It may be that for these displays that additional time was needed to recognize whether or not the displays were indeed blinking or not before the entire display was comprehended. Possibly speeding up the blinking rate would allow for faster recognition, but may also incur a distraction and/or annoyance cost.

Finally, it was surprising that significant differences were found for priority and descriptor across reaction time measures and accuracy. The effect was also independent of display. Thus, it may be that the message itself caused the effect. In other words, results showed that subjects were slowest at responding to medium priority messages indicating that medium priority is not very "interesting" or informative. Subjects may prefer only two levels of priority – low and high. The same could be true for descriptor in which the email descriptor was responded to faster than reminder or news.

6 Conclusions and Future Work

In summary, there are several conclusions that can be made from this study:

1. **There is a tradeoff between performance, preference, and display size.** As the number of LEDs decreased, accuracy, speed, and favorable opinion generally decreased. But, performance for the message set used here seemed to plateau at a size limit of four LEDs.
2. **This tradeoff can be mitigated by effective display design.** Using color, blinking, or binary codes appropriately allows the use of fewer LEDs to convey a fixed amount of information.
3. **Position provides the most intuitive representation.** Although position seems to aid message recognition the best, it also has the highest cost in terms of display size.
4. **Multi-color LEDs are useful for conveying more information in a smaller space.** Color was useful for distinguishing between three different types of messages and helps to reduce the number of LED's needed in a display but may not be helpful in conveying a large number of categories.

5. **Blinking may be better for attention getting than providing information.** Performance was not good with the blinking displays and other work [e.g., 2] has shown that few levels of blinking are useful, so blinking may not be a good choice for important meta-information.
6. **The mixing of certain visual factors may degrade performance.** There may be compounding or other effects caused by the mixing of certain visual representations. This is seen by comparing display 3, which had four LEDs but used both a binary representation and blinking, to display 4, which had only three LEDs but used only blinking. Blinking seemed to be harder to comprehend than binary coding, but having both at the same time made cue comprehension even more difficult.

It must be remembered, however, that these conclusions are being made based on the results of this limited, initial study, which is a first step in what is hoped to be a series of related work and additional experiments. This study has looked at a relatively small set of the total number of visual displays available, and there are many alternate implementations of each display. For example, information can be presented using different physical arrangements of LEDs and different binary codes. Learning curves were not investigated as part of this study, and no attempt was made to look at how the results might change over time. The LED displays were also simulated in a laboratory environment, rather than tested in real-world settings using real lights.

Overall, though, it appears that low-information-rate displays can be useful for providing information awareness in mobile situations. With the right display design, the performance reducing and experience spoiling effects of complex mappings can be mitigated. At the same time, this work supports the idea that even the smallest form factors such as watches, rings, and other jewelry can be used as valuable notification systems. This work opens the door for the design of a wide range of low-information-rate notification systems using a variety of information channels including other visual cues (e.g., motion and brightness), tactile information (e.g., vibration and pressure), and auditory information to name a few. Some issues that will be addressed in future work will be (a) measuring the distraction (individual and public) cost of low-information-rate displays in mobile situations, (b) determining what types of messages provide "critical bits" given a certain context or situation, (c) finding the number of messages or level of complexity at which performance breaks down to determine the "upper-bound" in information space for low-information-rate displays.

Other open research issues regarding notification cues include the analysis of end-user requirements and the need to determine exactly what types of information or events that users want to be notified of. The framework for mobile information needs presented in [22] begins to looks at this issue, but much more work is needed. There is also the issue of cue personalization. Allowing someone to customize the form that their notification cue takes may help alleviate some concerns over privacy and security of information, especially if the cue is sent in a public space. A final issue is that the acceptability of a notification cue or device, as with any sort of technology, may depend on fashion or social perceptions. Research in the area of wearable computing will undoubtedly influence the design of notification devices as well.

The results of this experiment are being integrated into the development of a notification cue prototype that alerts the user to approaching events. The events are recorded on the user's PDA, along with a priority for each event. The PDA vibrates when the event comes within a certain amount of time (e.g., half-an-hour) of the current time. A signal is also sent wirelessly to a set of LEDs that form a notification cue conveying a code for the priority of the event and how much time is left until the event. The notification is triggered based on the priority of the event. The PDA can be kept somewhere on the user's body (e.g., in a pocket), while the LEDs can be worn or kept elsewhere. The prototype is therefore provides a multi-modal notification cue (tactile and visual) that is subtle yet public. This work is an extension of [3].

References

1. Czerwinski, M., Cutrell, E. and Horvitz, E. (2000). Instant Messaging: Effects of Relevance and Timing. In S. Turner, P. Turner (eds), People and Computers XIV: Proceedings of HCI 2000, Vol. 2, British Computer Society, p. 71-76.
2. Grether, W. F., and Baker, C. A. (1972). Visual Presentation of Information. In H.A. Van-Cott and R.G. Kinkade (eds.), Human Engineering Guide to Equipment Design. Washington, DC. U.S. Government Printing Office.
3. Hansson, R. and Ljungstrand, P. (2000). The Reminder Bracelet: Subtle Notification Cues for Mobile Devices. In Extended Abstracts of CHI 2000, 323-324.
4. Hansson, R., Ljungstrand, P., and Redstöm, J. (2001). Subtle and Public Notification Cues for Mobile Devices. In Proceedings of UbiComp 2001.
5. Harris, L. R. and Jenkin, M. (2001). Vision and Attention. In M. Jenkin and L. R. Harris (eds.), Vision and Attention, Springer-Verlag, New York, 1-17.
6. Helander, M. G., (1987). Equipment and Workplace Design. In Salvendy, G. (ed.), Handbook of Human Factors, 1st Edition, (pp. 505-620). New York: Wiley-Interscience.
7. Holland, S. and Morse, D. R. (2001). Audio GPS: Spatial Audio in a Minimal Attention Interface. In M. D. Dunlop and S. A. Brewster (eds.), Proceedings of Mobile HCI 2001: Third International Workshop on Human-Computer Interaction with Mobile Devices.
8. Holmquist, L. E., Falk, J. and Wigström, J. (1999). Supporting Group Collaboration with Inter-Personal Awareness Devices. Personal Technologies, 3(1-2), 13-21.
9. Horvitz, E., Jacobs, A., and Hovel, D. (1999). Attention-Sensitive Alerting. In Proceedings of UAI '99, (Stockholm, July 1999) Morgan Kaufmann, 305-313.
10. Huang, E. M., Tullio, J., Costa, T. J., and McCarthy, J. F. (2002). Promoting Awareness of Work Activities Through Peripheral Displays. In Proceedings of CHI 2002, 648-649.
11. Johnson, P. (1998). Usability and Mobility; Interactions on the Move. In C. Johnson (ed.), Proceedings of the First Workshop on Human Computer Interaction for Mobile Devices. Scotland: University of Glasgow.
12. Kahneman, D., and Henik, A. (1981). Perceptual Organization and Attention. In M. Kubovy and J. R. Pomerantz (eds.) Perceptual Organization, Lawrence Erlbaum, Hillsdale, N.J., 1981, 181-211.
13. Mackinlay, J. (1986). Automating the Design of Graphical Presentations of Relational Information. ACM Transactions on Graphics, 5(2), April 1986, 110-140.
14. Maglio, P. P. and Campbell, C. S. (2000). Tradeoffs in Displaying Peripheral Information. In Proceedings of CHI 2000, 241-248.

15. McCormick, P. A. (1997). Orienting Without Awareness. Journal of Experimental Psychology. Human Perception and Performance, 23, 168-180.
16. McCrickard, D. S., Catrambone, R., Chewar, C. M., and Stasko, J. T. (2002). Establishing Tradeoffs that Leverage Attention for Utility: Empirically Evaluating Information Display in Notification Systems. International Journal of Human-Computer Studies, in press.
17. Miller, G. A., (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. The Psychological Review, Vol. 63, No. 2, March 1956, 81-97.
18. Moray, N. (1959). Attention in a Dichotic Listening: Affective Cues and the Influence of Instructions. Quarterly Journal of Experimental Psychology, 11, 56-60.
19. Sawhney and Schmandt (2000). Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments. ACM Transactions on Computer-Human Interaction, 7(3), 353-383.
20. Schmidt, A., Aidoo, K. A., Takaluoma, A., Tuomela, U., Van Laerhoven, K., and Van de Velde, W. (1999). Advanced Interaction in Context. In Gellersen, H.-W. (ed.) Proceedings of HUC '99, Berlin: Springer-Verlag, 89-100.
21. Tarasewich, P. (2003). The Design of Mobile Commerce Applications: What's Context Got to Do With It? Communications of the ACM, forthcoming.
22. Tarasewich, P. and Dideles, M. (2003). Towards a Framework for Information Needs in Mobile Environments. In F. Partovi (ed.), Proceedings of the Northeast Decision Sciences Institute 32nd Annual Meeting, 216-218.
23. Treisman, A. (1960). Contextual Cues in Dichotic Listening. Quarterly Journal of Experimental Psychology, 12, 242-248.
24. Van Dantzich, M., Robbins, D., Horvitz, E., Czerwinski, M. (2002). Scope: Providing Awareness of Multiple Notifications at a Glance. In Proceedings of AVI 2002.
25. Ware, C. and Beatty, J. C. (1985). Using Colour as a Tool in Discrete Data Analysis. Technical Report CS-85-21, Computer Science Dept., Univ. of Waterloo, Waterloo, Ont., Canada, Aug. 1985.
26. Wickens, C. D. and Carswell, C. M., (1997). Information Processing. In Salvendy, G. (ed.), Handbook of Human Factors and Ergonomics, 2nd Edition, New York: Wiley Interscience, pp. 89-129.
27. Wisneski (1999). The Design of Personal Ambient Displays. M.Sc. Thesis, May 1999, MIT Media Laboratory, Cambridge, MA.