What Can You Say With Only Three Pixels?

Christopher Campbell[1], Peter Tarasewich[2]

[1] IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120 USA
ccampbel@almaden.ibm.com
[2] College of Computer and Info. Science, Northeastern University, Boston, MA 02115 USA
tarase@ccs.neu.edu

Abstract. The size limitations of mobile devices can make information display especially difficult. Micro-displays must take into account the viability of different sizes and configurations for informing users, the flexibility they provide for different types of messages, and under which conditions these results are achieved. An experiment was performed to measure user learning and comprehension of five sets of messages of increasing information size and complexity on a simulated three-light visual display. Results show that these "pixel-based" displays can transmit detailed, information-rich messages up to 6.75 bits in size with minimal training.

1 Introduction

Computer devices and systems 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. In such technology-laden environments, information overload can be a serious problem. 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. 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 every time one becomes available. It may be too distracting (or 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.

To be successful, notification systems (and cues) must "present potentially disruptive information in an efficient and effective manner to enable appropriate reaction and comprehension" [10]. 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

To be presented at Mobile HCI 2004 September 13-16, Glagow, Scotland.
Proceedings published by Springer –Verlag.

can become quite complex, requiring the selection of appropriate delivery channels based on continuously changing contexts and dynamic information needs [9].

This paper presents results from an experiment that measured the learning and comprehension of visual notification cues that conveyed increasing amounts of information to the user. Three lights were used, each with three colors and two intensity levels. A previous study [14] showed that the three-light design (compared to 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.

2 Background

Notification cues can be visual, auditory, tactile, or multimodal in nature. They can be private such that only the receiver is aware of them, 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 [1].

The design and use of notification cues must take into account the intricacies of human attention, which involves the allocation of perceptual or cognitive resources to something at the expense of not allocating them to something else [2]. Humans have a limited amount of resources available for allocation to different tasks and cannot attend to everything at once. People can attend to a modality (e.g., vision, hearing, touch), a color, a shape, or a location [2]. The decision to attend specifically to one of these over the others arises from the task at hand.

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 and email notifications), the user’s environment outside of the computer is fairly consistent from day to day. Most offices and homes function with a fair amount of regularity and predictability, even if they experience a lot of activity. The user can devote relatively consistent attention to performing tasks on the computer.

On the other hand, 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 computer itself [13]. Furthermore, since devices are completely mobile, this outside environment can change rapidly. A mobile application may not be the focal point of the user’s current activities [3], as the user may be trying to balance interaction with a mobile device with other elements in the environment (e.g., walking along 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 a user can give to a mobile application will vary over time, and a user’s priorities can also change unpredictably [5].

An environment that consists of too many distractions can be confusing and unmanageable. Notification cues must be designed 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 ignored completely.

Much 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, current news headlines may scroll across a one-line display on a computer screen. Studies have looked at the effectiveness of presenting information in secondary displays in various formats [8, 10]. Research has generally found that performance on primary tasks is negatively impacted by secondary tasks [8], with some exceptions [10].

Other research has investigated notification systems and devices specifically for mobile environments. Wisneski [15] described a subtle and private notification device in the form of a watch that changes temperature as stock prices change. Holmquist, Falk, and Wigström [4] 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.

In a first attempt at creating a subtle notification cue that was also public, Hansson and Ljungstrand [1] 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 red LEDs that are triggered progressively as an event draws closer. With this device, people near the user can clearly see that the user is being notified about something.

Tarasewich et al. [14] conducted a study that measured the performance/size tradeoff of visual displays that ranged in size from two lights to nine lights, and used display characteristics, such as color and blinking, in various combinations. Results showed a reliable tradeoff between performance (response time and accuracy) and display size (number of lights). However, even the full set of twenty-seven messages used in the study could be conveyed with high recognition accuracy using only three lights by mapping the messages into color and position. The authors concluded that mobile devices with micro-level form factors could be designed to convey critical information and provide effective notifications. However, two issues were not explored in this study. One issue was how learning affected the comprehension and use of the visual displays. The second was how much information could be effectively conveyed using a display of a given size.

3 Evaluating Increasing Information Amounts

Mobile notification display designs 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., 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 for smaller and even micro level form factors (e.g., jewelry), and (c) using less power. Simply speaking, the less information conveyed, the less attention required to use it. However, less information does not mean the message is not informative. Even small amounts of critical information can be highly informative in mobile situations.

Table 1. Cue categories and associated values

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

An experiment was conducted to test comprehension and learning of visual cues conveying increasing amounts of information on a three-light display. [14] showed that a three-light design has a balance of good user performance and high user preference, all within a relatively small footprint. The mappings were chosen to be as simple and direct as possible using color and/or position for the information categories and values. The goals of the present experiment 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 on that display.

3.1 Information Mapping Functions

The same physical display size was used for each round of the experiment. Each simulated light could show the colors red, blue, and green. Each color could also display at one of two intensity levels – low or high (i.e., dim or bright). This means that we could theoretically encode six pieces of information on a single light (3 colors x 2 intensity levels). With three lights, we can encode a maximum of 216 (6 x 6 x 6) different messages. For this experiment, we chose five message sizes to display on the cue using five different mappings. The messages, based on one or more categories from Table 1, were mapped into the cue display using position, color, and intensity. Each cue represented information about a message that was available to the user.

Mapping 1. Here, all three lights were lit with the same high-intensity color. The color represented the source of the message; red for family, blue for friends, and green for work. This mapping used three lights to represent three messages.

Mapping 2. The three lights were the same used in Mapping 1. This time, however, the intensity of the color also varied. High intensity for a given color indicated that the message was an email. Low intensity indicated a voicemail. For example, if the lights were high intensity blue, this indicated an email message from friends. This mapping used the three lights to represent a total of six (3x2) messages to the user.

Mapping 3. The three lights were used, but each with three high-intensity colors (red, blue, green). The left light indicated source, the center type (new, reply, forwarded), and the rightmost 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. The lights represented twenty-seven (3x3x3) messages.

Mapping 4. The three 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 represented 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 represented 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 the possible alternatives. This is represented by:

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

where N is the number of alternative messages in the message-set [12]. Information loads range from 1.58 to 6.75 bits for message-sets 1 through 5 (see Table 2).

Display dimensions were chosen to produce intuitive mappings. For example, color is an especially effective way of coding nominal sets [7]. Kahneman and Henik [6] pointed out that the human visual system is effective at distinguishing a small number of distinct colors, so the three colors red, blue, and green were used. In addition, two easily discernable intensity levels were used for each color.

Table 2. Information load for each message-set

Message-Set	Alternatives	H (bits)
1	3	1.58
2	6	2.58
3	27	4.75
4	54	5.75
5	108	6.75

The message-sets are essentially notifications containing meta-data about messages for the user. These messages-sets were designed to facilitate learning and improve the performance of a three pixel display by (a) organizing messages by categories of meta-data, (b) mapping these categories directly to cues, and (c) providing for a progressive learning style. Categories provide subjects with a way to create mental “chunks” of information, thereby reducing the work of identifying messages. Without chunking, each message in the set of alternatives would have to be remembered, but with chunking, only the alternatives of a set of categories need be identified. Learning is further facilitated by providing a simple, direct, and consistent mapping between information and cues. In this case, the categories are mapped directly to visual cues. Finally, each larger message-set contains the mapping of the previous sets so that learning can be progressive, or built-up over time.

4 Methodology

Fifty-two undergraduate and graduate college students participated in this study. Eight subjects were female and forty-four were male. Ages ranged from eighteen to thirty years with an average age of twenty-three. None reported themselves as colorblind. Subjects were required to learn each message-set to a criterion level in order from the easiest (set 1) to the most difficult (set 5). They were only allowed to advance to the next message-set after achieving 90% correct on the current set.

The experiment was then conducted on a Pentium-4 computer running Windows XP with a 1024 x 768screen resolution. A Java program presented the cue displays as Graphics Interchange Format (GIF) files on a black background with a taskbar at the bottom. The GIF files were 555x250 pixels, with 96 pixels/inch resolution and 8-bit color depth. Status indicators showed the elapsed trial time and the number of correct answers for the current session. See Figure 1 for a sample program screen.

4.1 Design

The design was a one-factor (message-set) repeated measures design with five levels (see Table 2). 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.

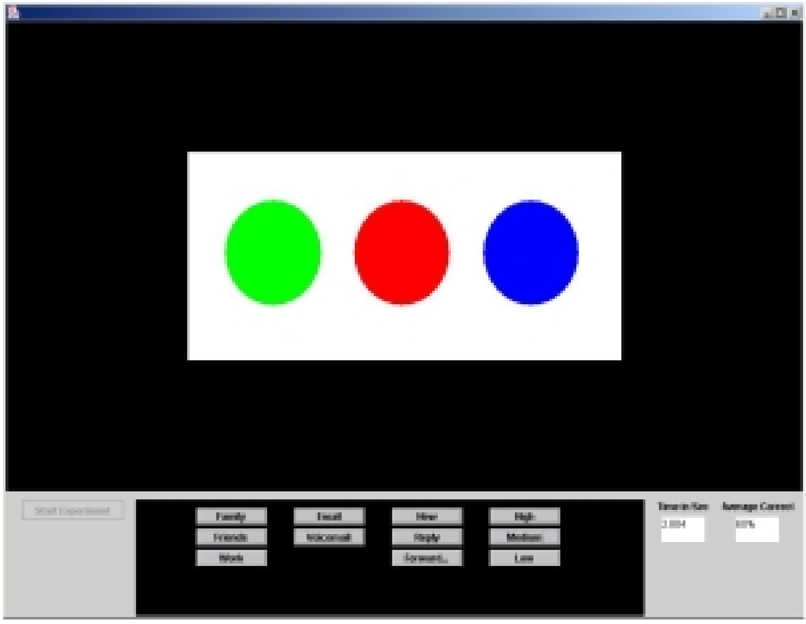

Fig. 1. Screen shot of testing environment

4.2 Procedure

Subjects first completed a questionnaire asking for background information. Each subject then completed five task sessions of increasing complexity that involved identifying notification cues. At the beginning of each session, a subject was shown a visual explanation of how information from a notification cue mapped to a specific visual display. The mappings started with Mapping 1 for the first session and finished with Mapping 5 for the last session. When a subject was ready to proceed, they were shown a notification cue. A subject responded by selecting one or more buttons on the screen corresponding to the information that was conveyed by the cue. A subject was then shown whether or not their response was correct, and they moved onto another cue in that session when they were 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 (otherwise, no payment was given). Each subject proceeded with the experiment at their own pace (except for the time-out), and could stop the experiment at any point. Response accuracies and task times were recorded.

During each session, buttons with the answer choices were listed in columns at the bottom of the screen. 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 was calculated after every answer on a moving basis over the ten most recent responses.

5 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

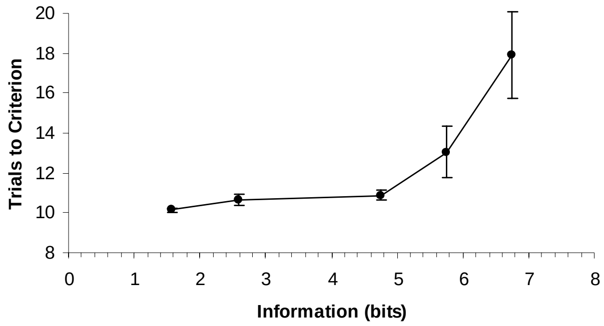

Fig. 2. Mean number of trials to reach criterion performance across information levels (±SE)

greater. Note that accuracy and number the trials are inversely correlated; the higher the accuracy, the fewer number of trials. 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.

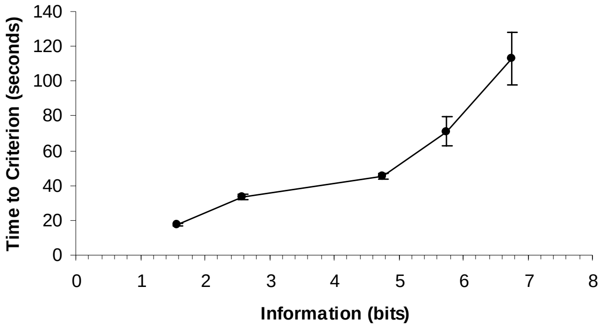

Fig. 3. Mean time to reach criterion performance across information levels (±SE)

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). As shown in Figure 2, the mean number of trials to criterion stays at ceiling performance (around 10 trials) for the first three message sets; but, significantly drops below the ceiling for message sets 4 (5.75 bits) and 5 (6.75 bits) (t(45) = 3.42, p < 0.001).

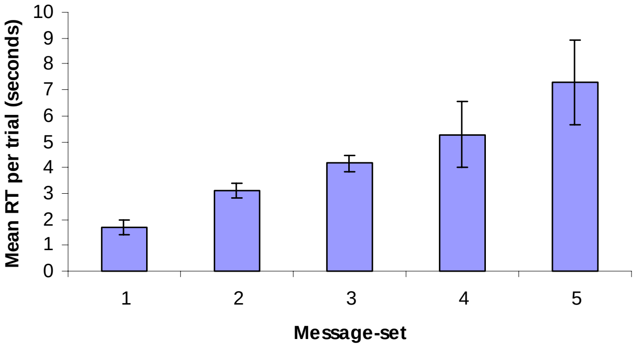

Fig. 4. Mean response time (RT) per trial across message-sets (±SE)

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). As Figure 3 shows, time to criterion increases steadily across information levels from a mean of 17 seconds to almost 120 seconds. Pairwise comparisons support this by showing reliable differences between information levels at 1.58 and 2.58 bits (t(45) = 10.7, p < 0.001);

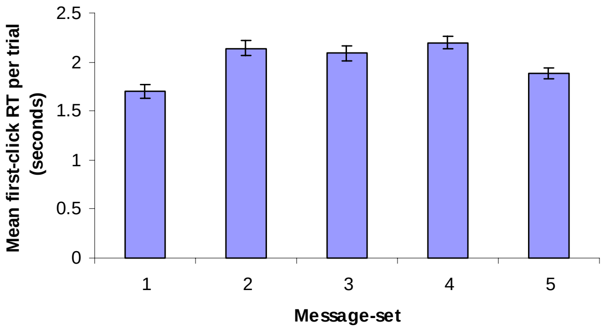

Fig. 5. Mean first-click response time (RT) per trial across message-sets (±SE)

2.58 and 4.75 bits (t(45) = 5.5, $p < 0.001$); 4.75 and 5.75 bits (t(45) = 3.2, $p = 0.0015$); 5.75 and 6.75 bits (t(45) = 2.9, $p = 0.0032$).

One difficulty with calculating the total time to criterion is that the increase in time could be the result, merely, of performing more trials with bigger message-sets. To examine this possibility, the mean response time per trial was calculated and is shown in Figure 4. This shows a reliable increase in response time per trial as the message-sets increase in terms of alternatives (F(4,180) = 502.85, $p < 0.001$). However, the increase is not as pronounced as in Figure 3.

One could further argue that the increase in response time for larger message-sets is solely the result of having to select more buttons in order to respond. Therefore, the mean first-click response time was analyzed (see Figure 5). This is calculated as the time from stimulus presentation until the first button is clicked. All these times are summed until criterion performance is reached and divided by the number of trials. A one way repeated measure ANOVA indicated a reliable increase in first-click response time across message-sets (F(4,180) = 17.16, $p < 0.001$). Subjects typically

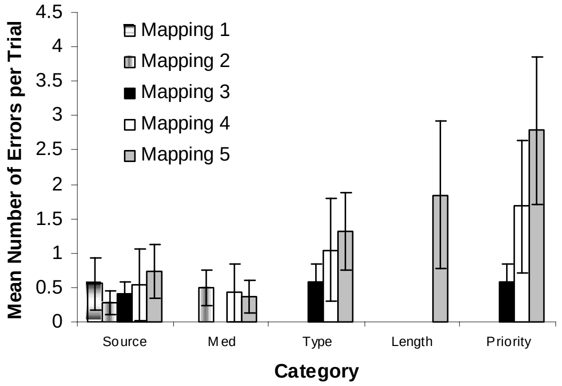

Fig. 6. Mean number of errors per trial across cue category for each mapping (±SE)

wait 2 seconds after the notification is displayed before responding except for message-sets 1 and 5.

Analysis of errors by cue category (see Table 1) shows approximately the same number of errors being made across all relevant mappings for Source and Medium (see Figure 6). There is, however, an increase in the number of errors for Type and Priority. This is shown as a significant increase in errors for Type and Priority over the remaining cue categories in Mappings 4 ($F(1,46) = 8.38$, $p = 0.006$) and 5 ($F(1,46) = 23.255$, $p < 0.001$).

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 for all subjects combined timed out and there was no reliable difference across message-sets ($F(4,180) = 1.71$, NS).

6 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, does start 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 (see Figure 3). This shows that there is a cognitive cost to learning 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 – we still see a significant increase in response time (see Figure 4). This assumes that subjects use a response strategy of first identifying all the cues and then responding. Several response strategies are possible:

1. Identify all cues and then respond.
2. Respond immediately with identified cues, identify remaining cues, then finish response.
3. Identify some cues, respond, identify more cues, respond, identify remaining cues, respond, etc.

It could be the case, for example, that when the number of response choices exceeds a certain amount, subjects switch to the second strategy listed. In other words, when faced with many buttons, subjects immediately click the buttons they know are correct, stop to consider the remaining choices, and finish responding. While this study does not definitively uncover which response strategy is used, the increasing first-click response time combined with the increasing trials to criterion suggest greater cognitive effort to identify notifications with more information.

Some support for strategies two or three comes from the analysis of errors by category (Figure 6). The number of errors should decrease for the same cue category repeated across Mappings due to learning. But subjects are probably reading and interpreting the cues from left to right, and the rightmost light receives the least attention. Even though Priority is interpreted exactly the same way in Mappings 3 and 4, Mapping 4 has an additional category to interpret on the leftmost light. This leaves less time to interpret the remaining lights because of the time limit, and results in more mistakes. This would also explain the increase in mistakes for Priority on Mapping 5 (along with the relatively high error rate for Length), because there is now an additional category to interpret on the rightmost light.

This suggests more important information should be placed on the “leftmost” part of a pixel-based notification display when the display is composed of elements aligned horizontally. Another potentially useful result is that, at least for the leftmost light of this three light cue design, there is no significant difference in error rates of the category mapped to color (Source) or to intensity (Medium).

7 Conclusions and Future Work

The goal of this study was to investigate the amount of information that could realistically be presented on pixel-based micro-sized displays. Results indicate that people can quickly learn fairly large notifications of over six bits with only three pixels. This makes low-information-rate, micro displays practical for people not willing to endure extensive training sessions. Design possibilities are also enhanced because many message schemes can be used 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 fact that stimulus-response (S-R) compatibility was high. S-R compatibility is the ease of transformation between the stimulus representation and the response [11]. In our case, the mapping between message categories (i.e., source, medium, type, length, priority) and response categories (i.e., set of buttons for each category) was direct. If, on the other hand, we had presented people with an array of buttons for each alternative message, performance would have degraded more quickly.

Another important factor was the organization of the message-sets into categories or chunks. For example, the 108 messages of 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. Chunking can also provide a variety of response strategies. As noted above, people can respond in a piecemeal fashion instead of all at once, effectively simplifying the identification task. Future work will compare message-sets that can be organized by category against those that cannot, and also against those for 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 hierarchical message-sets allows people to progressively learn instead of trying to memorize all messages at once. This appears to increase the size of messages that can be conveyed and reduce learning requirements. It also has the benefits of:

- Allowing for immediate use of the notification system with almost no learning

- Providing advanced functionality for advanced users while still offering simpler functionality for other users.

Future work will compare methods for designing notifications or messages-sets. 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. Other work will look at how customized information-cue mappings affect notification ease of use. Smaller form factors (e.g., watches) and other problem domains will also be explored.

There are many practical benefits to using pixel-sized visual notification cues. Theoretically, small lights (e.g., LEDs) can be embedded in almost any device or product. The cues can be sent quietly. They can also be customized to address privacy and security concerns. For example, three blue lights on a person's ring, even when noticed by other people, could convey a message only understood by the wearer.

References

1. Hansson, R., Ljungstrand, P., and Redstöm, J. Subtle and Public Notification Cues for Mobile Devices. *Proc. UbiComp 2001*, Springer-Verlag (2001).
2. Harris, L. R. and Jenkin, M. Vision and Attention. In M. Jenkin and L. R. Harris (Eds.), *Vision and Attention*, Springer-Verlag (2001), 1-17.
3. Holland, S. and Morse, D. R. Audio GPS: Spatial Audio in a Minimal Attention Interface. *Proc. Mobile HCI 2001*.
4. Holmquist, L. E., Falk, J. and Wigström, J. Supporting Group Collaboration with Inter-Personal Awareness Devices. *Personal Technologies*, 3(1-2), 13-21, 1999.
5. Johnson, P. Usability and Mobility; Interactions on the Move. *Proc. 1st Workshop on HCI for Mobile Devices*, Univ. of Glasgow (1998).
6. Kahneman, D., and Henik, A. Perceptual Organization and Attention. In M. Kubovy and J. R. Pomerantz (eds.) *Perceptual Organization*, Lawrence Erlbaum (1981), 181-211.
7. Mackinlay, J. Automating the Design of Graphical Presentations of Relational Information. *ACM Trans. on Graphics*, 5(2), 110-140, 1986.
8. Maglio, P. P. and Campbell, C. S. Tradeoffs in Displaying Peripheral Information. *Proc. CHI 2000*, ACM Press (2000), 241-248.
9. McCrickard, D. S., and Chewar, C. M. User Goals and Attention Costs. *Comm. of the ACM*, 46(3), 67-72, 2003.
10. McCrickard, D. S., Catrambone, R., Chewar, C. M., and Stasko, J. T. Establishing Tradeoffs that Leverage Attention for Utility: Empirically Evaluating Information Display in Notification Systems. *Intl. Journal of Human-Computer Studies*, 8, 547-582, 2003.
11. Morin, R. E. and Grant, D. A. Learning and Performance of a Key-Pressing Task as a Function of the Degree of Spatial Stimulus-Response Correspondence, *Journal of Experimental Psychology*, 49, 39-47, 1955.
12. Shannon, C. E. & Weaver, W. *The Mathematical Theory of Communication*. University of Illinois Press (1949).
13. Tarasewich, P. Designing Mobile Commerce Applications. *Comm. of the ACM*, 46(12), 57-50, 2003.
14. 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.
15. Wisneski, C. A. *The Design of Personal Ambient Displays*. M.Sc. Thesis, May 1999, MIT Media Laboratory, Cambridge, MA.