
Evaluating User Preferences for Adaptive Reminding

Julie S. Weber

Computer Science & Engineering
University of Michigan
Ann Arbor, MI 48109 USA
weberjs@umich.edu

Martha E. Pollack

School of Information
University of Michigan
Ann Arbor, MI 48109
pollackm@umich.edu

Abstract

We are developing an adaptive reminding system that tailors its reminders to its users' reminding preferences through real-time interaction and feedback. To determine the potential utility of such a system, we conducted a multi-phase user study, presented in this paper, in which we evaluate people's preferences for the visual presentation of reminders. Results indicate that people generally agree on the relative annoyance levels of visual reminders, and further, in certain contexts, more "annoying" or intrusive reminder styles are preferred. However, while there are some overarching patterns of agreement about the contexts in which certain types of reminders are preferable, preliminary evaluation also indicates that there are significant differences among people's preferences for specific visual reminders. This motivates the design and development of adaptive reminding systems that learn their users' individual preferences.

Keywords

Adaptation, evaluation, intelligent user interfaces, reminder systems, user preferences, user study

ACM Classification Keywords

H5.2. Information interfaces and presentation: User Interfaces—*Evaluation/Methodology, Interaction styles*

Introduction

Intelligent personal assistants are the focus of much current research. To date, these systems are somewhat rigid: they provide fixed assistance with scheduled tasks and activities, often overlooking the context in which this assistance is provided. We hypothesize that people have interaction preferences that affect their level of satisfaction with such systems. We further suggest that these preferences may be highly individualized and sensitive to the accompanying context, i.e., certain system behaviors may be more acceptable in some situations than others.

This research explores these hypotheses in the context of reminder systems. Specifically, we conducted experiments aimed at investigating the claims that (a) people will consistently classify certain types of reminder presentation styles as more or less intrusive, and (b) people will have very individualized preferences regarding how intrusive reminders should be when delivered within a given context. This includes both a user's current task at the time a reminder is delivered, and the importance and/or urgency of the task that is the object of the reminder.

To address the above hypotheses, we conducted a four-phase user study that evaluated reminding preferences with and in the absence of various combinations of contextual information. This paper will focus on the first and final phases of our study, which highlight and are representative of our primary findings. The first phase of the study asked participants to rate a set of eight visual reminder presentations by their relative acceptability. In the final phase of the study, contextual task and reminder utility information were combined in a set of eight scenarios comprising all

combinations of high and low importance, urgency, and degree of attentional focus associated with the current task. The initial studies that are described in this paper were preparatory to a larger study, currently underway, which we are conducting with "live" reminders generated in real-time from the actual calendars of office workers.

Background

Intelligent Personal Assistants

Various systems have been developed to provide time management assistance. Intelligent cognitive assistants provide their users with reminders about their daily activities and/or the steps within them. Some of these systems, such as Autominder [10] and COACH [2], use reasoning techniques to ensure that the reminders are adapted to the specific tasks that an individual must perform and the state of performance of those tasks. However, they do not explicitly reason about the user's preferences regarding the form or timing of reminders.

Reminder systems for use in the workplace include Cybreminder [3], Forget-Me-Not [7] and CALO [9], while Vastenberg et al [12] describe a notification system for use at home. While Cybreminder focuses on *when* to remind, and Forget-Me-Not compiles all of the information that could be required regarding *what* to remind, in neither case do these systems tailor their reminders to individual users' preferences, to specifically answer the question of *how* to remind. The same holds for Vastenberg's system for correlating notifications with their associated urgency levels.

CALO provides a broad range of user services, and the current project, which is part of the CALO effort, is a

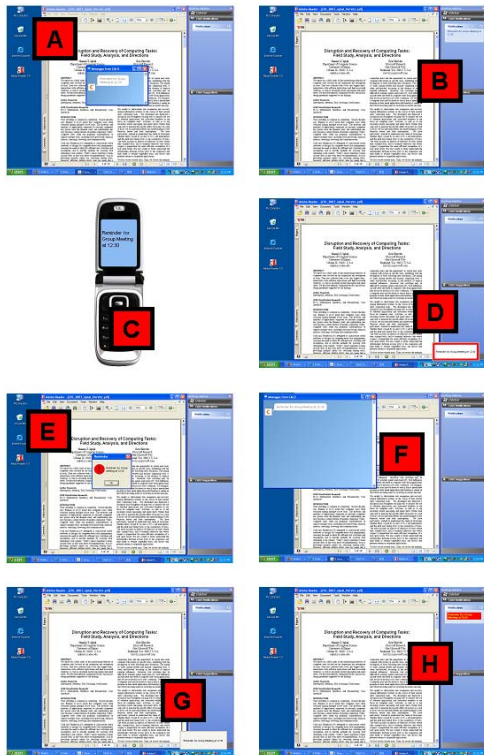


Figure 1: Each of the eight reminder presentation styles. Style *A* is a centered dialogue window, style *B* is a message on the persistent sidebar, style *C* is a text message on a mobile phone, style *D* is a newly opened message window enhanced with red, style *E* is a centered, modal, red dialogue, style *F* is a dialogue window in the upper corner, style *G* is a newly opened message window, and style *H* is a message on the sidebar in red.

first step in enabling CALO to tailor calendar-based reminders to its users' individual preferences.

Adaptive Interaction

Beyond reminder systems, there has been prior work on adaptive interfaces and interaction strategies. Examples of the former, adaptive interfaces, include [4, 1], which explore the question of how to provide a visual interface that is most desirable for a given user. Examples of the latter, adaptive interaction strategies, include [5, 6, 8] and focus on the questions of how and when to interrupt users, taking into account the tasks that they are currently performing.

Our work is closely related to both of these prior lines of research, in that we are interested in studying both how and when to interrupt computer users with reminders. However, a central conclusion of our research is that an approach based on associating specific interaction strategies with particular contexts across all users will not be sufficient: initial findings suggest that different users have different preferences, and hence systems must be designed to learn and adapt to individual users.

Study Design

We first describe an initial user study in which we investigated user preferences for visual reminders in terms of annoyance, or intrusiveness. In the following sections, we describe an additional phase of the study incorporating contextual task and reminder content information.

Participants

We had a total of 20 participants (15 male, 5 female) between the ages of 18 and 55. All participants in this

study were computer scientists and/or engineers comprising a sizable subset of the target user population of the CALO system for which our study was designed, and all but five had some degree of experience (as a user or developer) with CALO itself.

Reminder Presentation Styles

We define a *reminder presentation style* to be the set of variables representing the window type, screen placement, and color of a visual reminder, instantiated in various ways.

Window Type represents the manner in which the reminder is displayed on the screen. This feature can be instantiated with one of five different values, indicating that the reminder is displayed through (1) a natural language dialogue window; (2) a newly opened window (either modal or non-modal); (3) a previously open window, such as a persistent sidebar; and (4) a text message on a mobile phone.

The second feature of a reminder presentation style is called **Screen Placement** and determines whether a reminder is displayed (1) from the top left corner of the screen; (2) from the bottom right corner of the screen; (3) in the center of the screen; (4) on the right side of the screen; or (5) on a mobile phone.

Finally, the third feature of a reminder presentation style is its **Color**: whether it includes the color red as an accent color for increased attentional draw.

We sampled eight reminder presentation alternatives that combine different sets of presentation characteristics. (The size of the space precluded our performing tests with all possible feature combinations,

so we chose a representative set of feasible designs, i.e., those readily implementable within CALO.)

User Study: Annoyance

Our primary hypothesis for the first phase of our study was *H1: that in the absence of (explicitly defined) contextual task and utility information, different users will ascribe similar levels of perceived intrusiveness to the same reminder presentation style*. From initial informal interviews with potential system users, we found that *annoyance* is the primary factor that determines whether or not the user of an intelligent personal assistant actually continues to interact with the system; this is consistent with the claims of [8]. Consequently, we evaluate interaction styles on a scale of user annoyance, which can serve as an indicator of the overall efficacy of visual reminders.

Protocol

In this first phase, all of the participants were asked to consider the eight screenshots of reminder presentation styles (see Figure 1) and rate each presentation style by the amount of disruptive annoyance that it would cause. (By *disruptive annoyance* we mean the degree of disruption in a user's current task; we contrast this with annoyance that would be caused by not noticing a reminder.) The annoyance scale was presented to each participant as a number line from 0 to 10 with integer values ticked, and ratings were permitted to fall anywhere on the line. Lower values represented lower levels of annoyance. Participants could examine all screenshots concurrently and evaluate them in whatever order they desired.

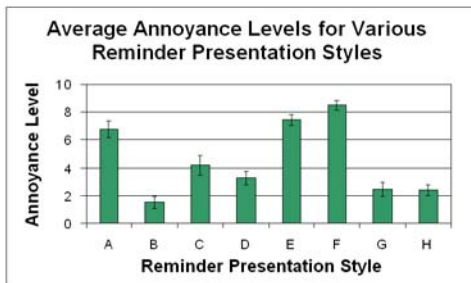


Figure 2: Average annoyance ratings among participants for each of the 8 visual reminders

Results

The mean annoyance values for each of the eight reminder presentation styles are displayed in Figure 2. Upon further analysis of these results using a repeated measures analysis of variance test and post-hoc pairwise evaluation with a Bonferroni adjustment, we found that instead of merely creating a continuous scale of increasing annoyance, the reminder presentation styles formed a pair of equivalence classes based on pairwise variations in annoyance values. (We also performed an identical analysis after normalizing the ratings to account for individual differences in rating distribution, and statistical results showed no significant differences.)

More specifically, reminder presentation styles *A*, *E* and *F* formed one equivalence class, in which relatively high levels of annoyance were ascribed, while *B*, *C*, *D*, *G* and *H* formed a second class with lower levels of ascribed annoyance. The members of the first class all had annoyance values that were statistically significantly different from all the members of the second class (using $\alpha=0.05$ throughout). Furthermore, within each of the two classes, none of the reminder presentation styles had statistically different annoyance values relative to one another with the exception of reminder presentation styles *B* and *C*. We label the two categories that are formed by the reminder presentation styles to be *Highly* and *Mildly* intrusive, respectively. An obvious question is why we adopt labels of intrusiveness rather than (perceived) annoyance. The answer is that a later phase of our study showed that when information about task context is included, users frequently reported a preference for the very same reminder types that they previously categorized as highly annoying. We therefore

1. You are trying to get through the end of a very mathematically rigorous technical document when a reminder arrives for a meeting with the Director of [your company] in five minutes. [HI,HU,HF]
2. You are sorting through old emails in your inbox when you are reminded about your meeting with the Director of [your company] in five minutes. [HI,HU,LF]
3. You are on an important teleconference in your office when you are reminded about next week's very relevant talk that you were hoping not to miss. [HI,LU,HF]
4. You are filling out what you find to be less than relevant paperwork when you're reminded about next Friday's application due date for the very prestigious award for which you're planning to apply. [HI,LU,LF]

Table 1a: Scenarios 1 through 4 (scenarios of high importance)

concluded that when a user reports a reminder presentation style as “annoying” in the absence of explicit contextual background, that user is actually describing the degree to which it intrudes on a current task.

User Study: Content & Task Context

Our primary hypothesis for this phase of the study is *H2: that there will be a hierarchy of scenarios with respect to their general acceptability ratings: scenarios with higher utility and lower task focus should be associated with a larger number of acceptable reminder presentation styles, and vice versa.*

Protocol

For this part of our study, each of a set of scenarios was read to a participant one at a time, and for each scenario the participant was asked to rate a reminder in each of the eight reminder presentation styles into one of three acceptability categories: *Best*, *Acceptable*, or *Unacceptable*. Because each scenario is meant to combine contextual task information with a reminder content-based utility value, each one includes the subject of the reminder and its start time or deadline, and it also identifies the task or activity that a participant is assumed to be performing at the time at which the reminder arrives. Scenarios are listed in Tables 1a and 1b. Our three variable measures, then, are task focus, reminder urgency and reminder importance), which we abbreviate “F,” “U” and “I,” and we use the letters “H” and “L” to indicate high or low values of each variable. Scenarios were read to the participants in random order to avoid data biases.

Each scenario represents a particular set of feature values. Because this was only a preliminary study,

each combination (e.g., high importance, low urgency, high attentional focus, abbreviated HI,LU,HF) was examined only once, to provide us with an initial understanding of people's preferences. Our current research effort involves examining such preferences in a true office environment in which participants will have the opportunity to rate each combination of reminding features numerous times.

Results

Results indicate that highly intrusive reminder presentation styles are much less acceptable, in general, than mildly intrusive presentation styles, regardless of reminder content and task context; this corroborates the results from the study's first phase. To specifically address hypothesis H2, which suggested that we would find a correlation between the number of acceptable reminder presentation styles and the level of utility and task focus associated with a given scenario, we performed a set of paired t-tests between every pair of the eight scenarios. We found (1) a significant difference in acceptability ratings between the single highest utility and low task focus scenario (HI,HU,LF) and all of the scenarios in which at least two utility/focus values were in their lower state; and (2) a significant difference between the single lowest utility and high task focus scenario (LI,LU,HF) and those four scenarios in which at least two utility/focus values were in their higher state. This provides preliminary evidence for a hierarchy affecting preferences, at least at the extremes. However, despite these general patterns, further evaluation of our results indicates that participants demonstrated highly individualized preferences for specific reminder presentation styles. Thus, a more detailed analysis of our resultant data will determine whether we can indeed consider each

5. You are on a very important teleconference in your office when, because you have not yet eaten lunch, you are reminded that the cafeteria will close in five minutes. [LI,HU,HF]
6. You are packing up and heading to another building when you are reminded that today is the last day to re-subscribe for your virus defense package before prices increase by five dollars. [LI,HU,LF]
7. You are in your office, meeting with your boss, when a reminder arrives for a meeting that you had no interest or intention of attending. [LI,LU,HF]
8. You are taking a break to stretch your legs when a reminder arrives for a talk that you know has been canceled. [LI,LU,LF]

Table 1b: Scenarios 5 through 8 (scenarios of low importance)

category of reminder (in terms of context & utility) independently.

Implications & Future Work

The objective of our initial study was to attempt to characterize people's preferences for visual reminders in the context of varying reminder utility and task focus. Measuring such preferences led to the preliminary conclusion that while there is general agreement on the relative intrusiveness of visual reminders, people also tend to have highly varied reminding preferences in different contextual situations. Of course, the use of artificial reminders outside of true task context does limit the generality of the results, but it is not uncommon to use an artificial context as a preliminary source of insights prior to full-scale user studies (e.g., [12]).

Our next step is to evaluate visual reminding styles in real-world contexts, associated with actual calendar events. Finally, the long-term goal of our work is not only to understand the space of user preferences, but to develop methods that enable a system to learn such preferences via interaction, and then to make use of them to create a better user experience.

References

- [1] Benyon, D. R. Accomodating Individual Differences Through and Adaptive User Interface. *Adaptive User Interfaces—Results and Prospects*, 1993.
- [2] Boger, J., Poupart, P., Hoey, J., Boutilier, C., Fernie, G. and Mihailidis, A. A Decision-Theoretic Approach to Task Assistance for Persons with Dementia. In *Proceedings of IJCAI*, 2005.
- [3] Dey, A. K. and Abowd, G. D. Cybreminder: A Context-Aware System for Supporting Reminders.

Second International Symposium on Handheld and Ubiquitous Computing, 1927 (2000) 172-226.

- [4] Gajos, K., and Weld, D. S. Preference Elicitation for Interface Optimization. In *Proceedings of UIST*, 2005.
- [5] Gluck, J., Bunt, A., and McGrenere, J. Matching Attentional Draw with Utility in Interruption. In *ACM CHI*, 2007.
- [6] Hudson, S. E., Fogarty, J., Atkeson, C. G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J., and Yang, J. Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. In *ACM CHI*, 2003.
- [7] Lamming, M., and Flynn, M. Forget-Me-Not: Intimate Computing in Support of Human Memory. In *Proceedings of UIST*, 2005.
- [8] Levitt, J. Internet Zone: Good Help is Hard to Find. Information Week: Listening Post 2001. <http://www.informationweek.com/835/35uwjl.htm>.
- [9] 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. *International Journal of Human-Computer Studies*, 58, 5 (2003), 547-582.
- [10] Myers, K., Berry, P., Blythe, J., Conley, K., Gervasio, M., McGuinness, D., Morley, D., Pfeffer, A., Pollack, M., and Tambe, M. An intelligent personal assistant for task and time management. *AI Magazine* 28, 2 (2007), 47-61.
- [11] Pollack, M. E., Brown, L., Colbry, D., McCarthy, C. E., Orosz, C., Peintner, B., Ramakrishnan, S., and Tsamardinos, I. Autominder: An Intelligent Cognitive Orthotic System for People with Memory Impairment. *Robotics & Autonomous Systems*, 44 (2003), 273-282.
- [12] Vastenburg, M. E., Keyson, D. V., and de Ridder, H. Considerate Home Notification Systems: A Field Study of Acceptability of Notifications in the Home. *Personal and Ubiquitous Computing*, June 2007.

Acknowledgments

We thank Neil Yorke-Smith for his assistance in the development our study. Thanks are also due to all of the individuals who took the time to participate in this study. We further acknowledge and greatly appreciate the support of SRI International.