Comprehensive User Evaluation of Adaptive Graphical User Interfaces

Leah Findlater and Joanna McGrenere
University of British Columbia
Vancouver, Canada
{lkf, joanna}@cs.ubc.ca

Abstract
Adaptive graphical user interfaces have the potential to improve the user's experience by personalizing the interface to better suit his or her needs. However, evaluations have yielded conflicting results: some studies show that adaptive menus and toolbars are faster or preferred in comparison to static counterparts [6,8], while others show the opposite [3,10,11]. We propose an evaluation approach that should allow for more comprehensive comparison of adaptive interfaces by: (1) reporting on and isolating factors such as accuracy, stability and predictability that may contribute to the user's experience, and (2) measuring the user's awareness of advanced features, in addition to performance and user satisfaction.

Keywords
Adaptive interfaces, personalization, evaluation.

CHI 2008, April 5 – April 10, 2008, Florence, Italy
Workshops and Courses: Usable Artificial Intelligence.

Introduction
Feature-rich interfaces graphical user interfaces (GUIs) can offer necessary functionality for many users but may be overwhelming for some users, especially novices. Adaptive GUIs have the potential to address this issue by personalizing menus, toolbars and other control structures to better suit an individual user's needs, thus reducing complexity. Although some commercial examples of adaptive GUIs exist, such as the menus in Microsoft Office versions 2000 to 2003, these approaches have yet to be widely adopted. Evaluations in the research literature have yielded conflicting results, showing that adaptive menus and toolbars may be faster or preferred in comparison to static counterparts [6,8], or, conversely, that static or adaptable (user-controlled) approaches are best [3,10,11].

These conflicting results are likely due to variation in characteristics between the adaptive approaches that have been studied; for example, adaptive accuracy, which is the percentage of time the system correctly predicts the next feature needed by the user, can affect user performance [5,6,7,13]. In addition to adaptive accuracy, recent work [2,5,7] has begun to explicitly incorporate predictability and stability of the interface into user evaluations. These three characteristics should be relatively straightforward to report alongside user evaluation results, and are applicable both to purely adaptive approaches, and to approaches that involve some degree of user control (mixed-initiative).

We propose a two-pronged approach to conduct more comprehensive user evaluations of adaptive and mixed-initiative GUIs. First, detailed reporting is needed of adaptive interface factors (e.g., accuracy, predictability, and stability), and these characteristics should be incorporated, where appropriate, into study designs. Second, the traditional measures of efficiency and satisfaction should be recorded in conjunction with *awareness*, a measure we have introduced to assess the impact that personalization may have on incidental learning of the full set of features in an interface [4]. The combination of these two steps should lead to more meaningful comparisons of adaptive GUIs.

Isolating Adaptive Interface Factors

Evaluations have begun to isolate the effect of adaptive accuracy, predictability, and stability on the user's experience. Results have shown that higher accuracy adaptive GUIs can make users faster and can result in greater utilization of adaptive predictions by the user [5,6,7,13]. However, the effectiveness of individual adaptive designs may interact with accuracy: Tsandilas and schraefel found that lower adaptive accuracy impacted the user's error rate, but that this was true for only one of two adaptive menu designs they studied [13].

In contrast to accuracy, stability and predictability have not been as well-studied, and the terminology requires clarification. *Stability* can be quantified as the frequency with which the adaptive algorithm causes features to change in the interface (e.g., moves or hides a feature). Both Cockburn, Gutwin and Greenberg [2], and Bridle and McCreath [1] have highlighted the importance of considering stability in evaluations of adaptive interfaces (although Bridle and McCreath called it predictability). In comparison, *predictability* may have a more subjective component. Gajos et al. define predictability as the ease with which users can understand and predict the behavior of the adaptive algorithm [7], which is also similar to the definition we have used previously [5]. In general, a more stable interface should be more predictable for the user; however, this is not necessarily the case and it may be useful to distinguish between the two for user evaluations. For example, adaptive changes based strictly on the set of most recently used features may be easily understood by the user, making the adaptive behaviour predictable (as shown in [7]), but depending on the specific sequence of features used, this approach may not necessarily offer high stability.

Our experience also emphasizes the importance of considering accuracy, stability, and predictability in combination. We conducted a controlled lab study to compare adaptive menus with two levels of accuracy (50% vs. 78%) to a static control condition [5]. The adaptive menus were implemented as split menus, where three adaptively predicted items were replicated at the top of the menu. Our goal was to create two adaptive conditions that were as similar as possible in respects other than accuracy, and we expected that users would prefer the higher accuracy condition. Pilot testing showed a surprising result though: when asked which condition they preferred, 2 out of 4 users reported that they found the low accuracy menus more predictable. To address this possible confound between accuracy and predictability in the full study, we modified the adaptive algorithm and specifically sought to create two equally predictable accuracy conditions.

Isolating one of accuracy, stability, or predictability as a factor in a study design can be challenging. For example, to achieve two levels of accuracy for adaptive item lists, Tsandilas and schraefel changed the set of adaptive predictions for each trial in their study, either including the item to be selected or not [13]. Lower accuracy conditions would have included more randomly highlighted items, which likely resulted in different levels of predictability. In a study of adaptive toolbars, Gajos et al. used two different experimental tasks that resulted in different levels of accuracy for the same interface [6]. Again, this did not purely isolate accuracy, because the tasks may have impacted other aspects of adaptive behaviour. In our work, we used an identical set of selections for each condition and determined adaptive predictions by: (1) applying a base algorithm, and (2) adjusting predictions to achieve a desired level of accuracy while adhering to several constraints (e.g., to maintain the level of predictability where possible) [5]. Even so, it was difficult to achieve equal stability for both accuracy conditions. Careful study design will be needed to control for these adaptive factors in future evaluations.

Awareness in Personalized Interfaces

Since personalization is a goal of many adaptive interfaces, the second component of our proposed evaluation approach is to assess both the negative and positive aspects of personalization. While drawbacks have been identified for both adaptable and adaptive personalization mechanisms [9,12], research on such approaches generally accepts that personalization itself is beneficial. One of the goals of personalization is to reduce or reorganize the set of features in the interface to increase efficiency; in doing so, the tradeoff may be a negative impact on the user's awareness of the full set of available features in the application. To evaluate this aspect of personalized interfaces, we proposed *awareness* as a measure of the incidental learning that occurs as the user works in an interface [4].

Experimental results have shown a measurable tradeoff between efficiency and awareness. In an initial study incorporating awareness, we compared two layered interface designs (an adaptable personalization approach) to a static control condition [4]. The minimal layered interface improved efficiency over the control condition, but subjects in the minimal layered condition were not as aware of advanced features after transitioning to the full interface as those who had worked in the full interface from the outset. More recently, in studying adaptive menus, we also found that adaptive accuracy impacts the user's awareness of advanced features [5]. The higher accuracy adaptive menus were more efficient, but because they better focused the user's attention on only a subset of features, they resulted in lower awareness of unused features than the lower accuracy menus. Reduced awareness of advanced features in the interface may reduce the user's longer-term learning and efficiency.

What is considered to be a desirable balance between efficiency and awareness may change in different design contexts. High awareness of advanced features may be more important for software applications where users are expected to mature into experts, for example, as with a complex integrated development environment. An adaptive mechanism that could predict new, potentially useful features for users may be beneficial in these situations. On the other hand, for applications that are used on a less frequent basis or for those applications that cater to a range of users with varying levels of expertise, short-term efficiency may be more important than awareness.

Conclusion

We have proposed that more comprehensive evaluations of adaptive GUIs may be achieved by isolating accuracy, stability, and predictability, and by measuring awareness in addition to efficiency and satisfaction. Our goal for this workshop is to encourage discussion on the strengths and weaknesses of the proposed approach, and to address the following:

- How should predictability and stability be operationalized?
- How can we design studies to isolate accuracy, predictability, and stability, and to determine the relationship between the three?
- How might the relative importance of accuracy, stability, and predictability, and the desirable balance between efficiency and awareness change in different application domains?

Providing more comprehensive reporting of adaptive interface factors and the application of efficiency, awareness, and user satisfaction should provide more meaningful comparisons among designs. It will be useful to understand, for example, whether stability and accuracy have equally significant impacts on user experience. More detailed reporting of adaptive characteristics may also help to clarify the previously conflicting results on adaptive GUIs.

References

[1] Bridle, R. and McCreath, E. Inducing shortcuts on a mobile phone interface. *Proc. IUI*, (2006), 327-329.

[2] Cockburn, A., Gutwin, C., and Greenberg, S. A predictive model of menu performance. *Proc. CHI '07*, (2007), 627-636.

[3] Findlater, L. and McGrenere, J. A comparison of static, adaptive and adaptable menus. *Proc. CHI 2004*, (2004), 89-96.

[4] Findlater, L. and McGrenere, J. Evaluating reduced-functionality interfaces according to feature findability and awareness. *Proc. IFIP Interact*, (2007), 592-605.

[5] Findlater, L. and McGrenere J. Impact of screen size on performance, awareness, and user satisfaction with adaptive graphical user interfaces. *Proc. CHI '08*, (2008). To appear.

[6] Gajos, K. Z., Czerwinski, M., Tan, D. S., and Weld, D. S. Exploring the design space for adaptive graphical user interfaces. *Proc. AVI '06*, (2006), 201-208.

[7] Gajos, K.Z., Everitt, K., Tan, D.S., Czerwinski, M., Weld, D.S. Predictability and accuracy in adaptive user interfaces. *Proc. CHI 2008*, (2008). To appear.

[8] Greenberg, S. and Witten, I. Adaptive personalized interfaces: A question of viability. *Behaviour and Information Technology* 4, 1(1985), 31-45.

[9] Höök, K. Steps to take before intelligent user interfaces become real. *Journal of Interacting with Computers* 12, 4(2000), 409–426.

[10] McGrenere, J., Baecker, R., and Booth, K. An evaluation of a multiple interface design solution for bloated software. *CHI Letters* 4, 1(2002), 163–170.

[11] Mitchell, J. and Shneiderman, B. (1989). Dynamic versus static menus: An exploratory comparison. *SIGCHI Bulletin* 20, 4(1989), 33-37.

[12] Shneiderman, B. and Maes, P. Direct manipulation vs. interface agents: Excerpts from debates at IUI 97 and CHI 97. *Interactions* 4, 6(1997), 42-61.

[13] Tsandilas, T. and schraefel, m. c. An empirical assessment of adaptation techniques. *CHI '05 Extended Abstracts*, (2005), 2009-2012.