Insights from the Design and Evaluation of a Mixed-Initiative Personalization Facility

Andrea Bunt
School of Computer Science
University of Waterloo
Waterloo, ON, Canada
abunt@cs.uwaterloo.ca

Cristina Conati
Department of Computer Science
University of British Columbia
Vancouver, BC, Canada
conati@cs.ubc.ca

Joanna McGrenere
Department of Computer Science
University of British Columbia
Vancouver, BC, Canada
joanna@cs.ubc.ca

CHI 2008, April 5 – April 10, 2008, Florence, Italy
ACM 978-1-60558-012-8/08/04.

Keywords
Personalization, mixed-initiative interactions, user control, explanations, evaluation

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
Software applications are often rich in functionality to accommodate users with diverse needs. Consequently, their GUIs can be complex, more so than is necessary from an individual user's perspective. One means of helping users cope with this complexity is to provide them with a GUI that is p*ersonalized* to their specific needs [7]. In this paper, we highlight issues from the design and evaluation of a specific approach to personalization: a mixed-initiative solution, where both the user and system participate in the personalization.

Mixed-initiative strategies [4] for GUI personalization combine aspects of: (i) *adaptive* approaches (e.g. [3]), which rely on AI techniques to personalize the GUI automatically; and (ii) *adaptable* approaches (e.g., [7]), which rely on users to personalize on their own through direct manipulation interface mechanisms. In combining elements of adaptive and adaptable approaches, the goal of a mixed-initiative system is to leverage each of their respective advantages while ideally minimizing their disadvantages (see [1] for a discussion of their advantages and disadvantages).

Our mixed-initiative approach to GUI personalization is as follows. First, we provide users with a direct manipulation facility that leaves them in control of deciding when and how to personalize. We augment this facility, however, with system-generated recommendations designed to increase personalization effectiveness and decrease personalization effort. A unique aspect of our approach is that these recommendations are dynamically generated based on a formal assessment of how different personalizations will impact the individual user's task time. This assessment considers user-specific traits, such expertise and anticipated feature usage -- information on which is stored in a user model.

We have implemented our approach within the MICA (**M**ixed-**I**nitiative **C**ustomization **A**ssistance) system and have applied the framework to MSWord. An overview of MICA and its two evaluations can be found in [1] and [2]. Here we discuss aspects of this work that apply to the development of usable AI. We begin with interface design, focusing on ways of maintaining control and interaction transparency/predictability, while still accommodating individual preferences. Next we discuss evaluation, concentrating on how we evaluated early in the design process.

Mixed-Initiative Interface Design

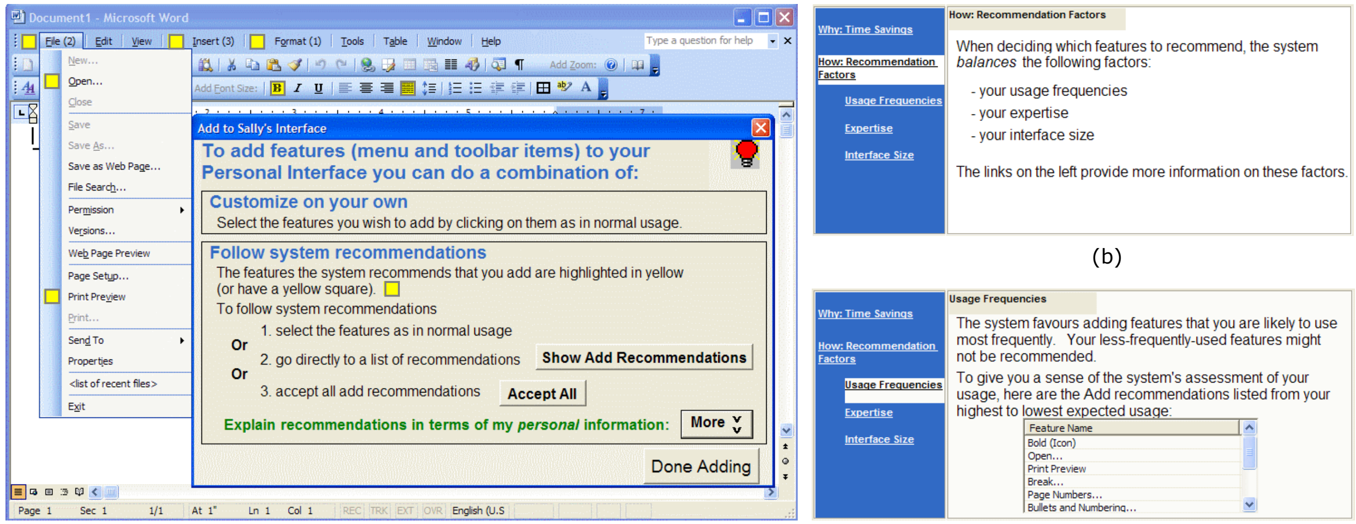
The literature (e.g., [5]) suggests that a key to usable AI is maintaining two properties: (i) user control and (ii) interaction transparency and predictability. This section discusses how we implemented these properties within MICA's mixed-initiative interface.

Multiple Levels of User Control

MICA maintains a high degree of user control by leaving the final decision of when and how to customize to the user. Similar to other studies (e.g., [6][7]), however, during our pilots and two formal evaluations, we observed individual differences in users' desire for control. Therefore, MICA's interface provides users with multiple ways to follow recommendations, each of which differs in its level of control afforded to the user. At one end of the spectrum, users can fully control personalization, with the recommendations simply acting as a visual guide (see the menus and toolbars in Fig. 1(a)). At the other end of the spectrum, users can delegate most of the responsibility to the system, by following all recommendations with a single button click ("Accept All" in Fig. 1(a)).

Interaction Transparency and Predictability

To increase interaction transparency and predictability, MICA explains the rationale underlying its adaptive behaviour (e.g., see Fig. 1 (b) and (c)). Similar to users' desire for control, in our second evaluation, we observed individual differences in users' desire to have the explanation facility present and its impact on user attitudes. The explanations increased feelings of trust, understanding and predictability for some users. Other users, however, found the interaction to be sufficiently transparent and predictable without the explanations, or didn't perceive these characteristics to be important



(a) (b) (c)

Figure 1. MICA's interface for adding features to the personalized interface, which pops up when the user enters the personalization facility (a similar interface exists for deleting features). The main dialogue box and recommendations (yellow squares/highlights in the menus/toolbars) are in (a), while (b) and (c) show parts of MICA's rationale component (accessible through the "More" button in (a)).

within this setting. To accommodate these different reactions, MICA's rationale component is accessible from the main mixed-initiative dialogue (see Fig. 1(a)), but not displayed until requested by the user.

The downside of MICA's multiple levels of control and explanation facility is that they increase the complexity of the personalization facility. Introducing usable AI with a minimal amount of accompanying complexity is a challenge for future work. Despite our use of an iterative design and evaluation process to formulate the rationale component and help ensure its clarity, there is still room for improvement in this area. For example, it may be possible to reduce the amount of text in the explanations and perhaps make them more graphical.

Evaluations

We have conducted two laboratory evaluations of MICA, both of which used a within-subjects design. The first compared MICA to an adaptable alternative, and indicated MICA has positive impacts on both task time and a number of qualitative measures [1]. The second evaluation compared versions of the system with and without the explanation facility [2]. Some of the key results from this study were highlighted in the previous section.

System Fidelity for User Testing

With research that combines AI and HCI, a key issue is how much of the system to implement prior to the first set of evaluations. On the one hand, developing a complete system that can capture, assess, and reason about relevant user characteristics has the potential to be an interesting AI contribution. On the other hand, early evaluation of a low-fidelity prototype can generate considerable insight into the advantages and disadvantages of the proposed adaptive behaviour, with far less implementation time.

In regards to this tension, we implemented MICA's general framework, which performs comprehensive performance-based reasoning to recommend suitable personalizations; however, we conducted our two evaluations without a fully functional user model. Recall that MICA's user model includes information on user expertise and anticipated feature usage, which MICA uses to make its recommendations. To validate the framework prior to implementing appropriate online assessment algorithms within the user model, we used a variant of the "Wizard of Oz" technique. Participants were given the impression of interacting with a fully functional system, however, user expertise and features usage were initialized in the user model by (a) administering detailed expertise questionnaires and (b) having users complete scripted tasks, which gave us fairly accurate information on feature usage.

The danger in evaluating at this stage is that the system might be less usable when working with the actual, likely less accurate, user model. One solution is to evaluate the system at multiple pre-determined levels of accuracy (e.g., [8]), but this can lead to an infeasible number of experimental conditions (particularly for a within-subjects design). If it is not feasible to evaluate at multiple levels of accuracy, ideally one would want to introduce "realistic" noise into the user model's assessments. Challenges with this approach include determining not only the type and frequency of system errors, but also how to make these errors seem plausible, even in a laboratory setting.

Summary and Future Work

In this paper we related our experience in designing and evaluating a mixed-initiative system for GUI personalization. Future work includes: (i) designing explanations that accurately describe complex adaptive behaviour, yet are still lightweight; and (ii) introducing realistic noise into "Wizard of Oz" evaluations.

References

[1] Bunt, A., Conati, C. and McGrenere, J. Supporting Interface Customization Using a Mixed-Initiative Approach. In *Proc. IUI*, (2007) 92-101.

[2] Bunt, A., McGrenere, J. and Conati, C. Understanding the Utility of Rationale in a Mixed-Initiative System for GUI Customization. In *Proc. UM*, (2007), 147-156.

[3] Gajos, K. and Weld, D.S. SUPPLE: Automatically generating user interfaces. In *Proc. IUI*, (2004), 93–100.

[4] Horvitz, E. Principles of mixed-initiative user interfaces. In *Proc. CHI*, (1999), 159–166.

[5] Jameson, A. Adaptive interfaces and agents. In Human-Computer Interaction Handbook, (2003), Erlbaum, 305–330.

[6] Jameson, A. and Schwarzkopf, E. Pros and cons of controllability: An empirical study. In *Proc. AH*, (2002), 193–202.

[7] McGrenere, J., Baecker, R.M., and Booth, K.S. An evaluation of a multiple interface design solution for bloated software. In *Proc. CHI*, (2002), 163-170.

[8] T. Tsandilas and m.c. Schraefel. An empirical assessment of adaptation techniques. *Ext. Abstracts CHI*, (2005), 2009–2012.