Designing Evaluations with High Ecological Validity

Matei Negulescu
University of Waterloo
David R. Cheriton School of Computer Science
mnegules@cs.uwaterloo.ca

Jaime Ruiz
University of Waterloo
David R. Cheriton School of Computer Science
jgruiz@cs.uwaterloo.ca

Edward Lank
University of Waterloo
David R. Cheriton School of Computer Science
lank@cs.uwaterloo.ca

CHI 2010, April 10–15, 2010. Atlanta, Georgia, USA
ACM 978-1-60558-930-5/10/04.

Abstract
Designing ecologically valid studies that accurately gauge the learnability, usability and overall efficiency is often difficult. Previous research on sketch based interaction techniques are often evaluated using simplified tasks. In this paper, we propose evaluating new pen/stylus interfaces using experiments with higher ecological validity. As an example, we present a study of a novel interaction technique – the inferred mode protocol. The study looks at learnability in the real world by considering multiple scenarios: users with varying level of training, and different ways of presenting computational intelligence. The study design helped extract previously undiscovered important information on the learnability and usability of the inferred mode protocol.

Keywords
inferred mode, sketch interface, pen, tablet, stylus

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User interfaces – Evaluation/ methodology

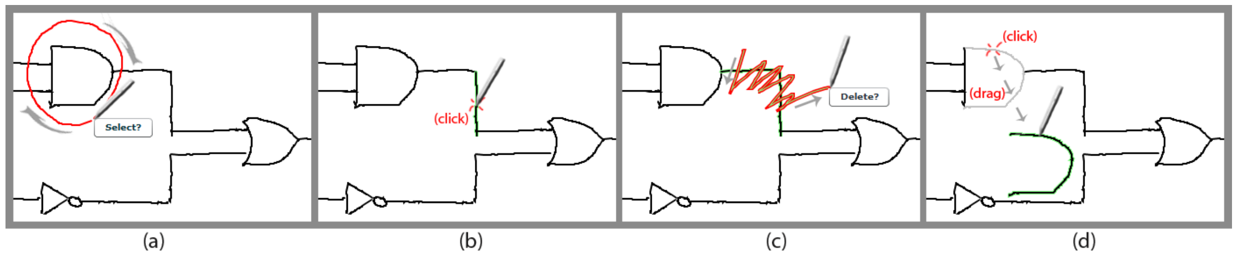

Figure 1: The inferred mode protocol. Panel a. shows smart circle select. When an object is circled, a mediator appears (top), but no mediator appears if the circle encloses nothing. Panel b. shows smart select click. Panel c. shows smart delete (top) and shading (bottom). Panel d. shows translation (top) vs smart drawing (bottom).

Introduction

The advent of tablet PCs has led to research into new interaction techniques that take advantage of the new affordances [2, 1]. Moreover, recent work by Ruiz et al. has shown that the cognitive load associated with tradiitional interfaces increase when adding more operational modes [4]. As a result, various alternatives to software modes in pen/tablet interfaces were considered, the most important to this paper being gesture based interfaces [5].

All of the techniques proposed to either simplify mode switching or to eliminate explicit modes from sketch interfaces have been evaluated experimentally. However, many of the evaluations performed have used simplified tasks like pie-cutting [2] or line drawing [4], or have used discrete command invocation evaluation [1] where the user is told to perform a specific command – 'delete', 'cut', 'copy', or 'select'. While these evaluations are useful in telling us about speed and error rate in controlled conditions, they tell us little about the usability of techniques in real-world drawing tasks.

When evaluating new interaction techniques, it is important to take into consideration real life scenarios and users. The previously mentioned evaluations are among those that make certain assumptions about their users. Firstly, the users are required to learn the technique before undertaking the task. This is problematic as users in the real-world are often poorly trained and adopt a "walk-up-and-use" approach to new software. Secondly, the users are forced to use the technique in order to gauge metrics such as efficiency and error rate [2, 4]. Though this approach provides good internal validity, many interfaces allow the users to achieve their goals using multiple paths. Lastly, the experiments often provide rigid tasks that do not relate well to real life use of the application. Taking into account realistic usage of the application can reveal useful information not occuring in artificial experiments.
As an example, we describe an experiment with high ecological validity that allowed us to reveal new usability benefits/problems of the inferred mode protocol that did not appear in the initial evaluation of the technique [5]. The inferred mode protocol allows the user to perform draw, select, and delete operations by interpreting new gestures based on the context. Figure 1 depicts the protocol's interaction paradigm.

Methodology

The goal of the study was to measure user adoption of the inferred mode protocol in a realistic sketching activity, both from the perspective of learnability and user preference. Over time, in an ecologically valid experiment, user preference can be measured by comparing the frequency of use of inferred mode features with the frequency of use of other options available in the interface. If participants use either inferred mode or alternatives more frequently, we can claim that there is a preference for one or the other.

To study learnability, we divided our participants into two groups, those who received instruction and those who did not. The participants in the Instructed group were also give a three minute overview of how the inferred mode protocol worked in the sketching interface they were using, while participants in the Not Instructed group were not. To limit bias, we were careful to show participants in the Instructed group both mechanisms for changing modes in the interface, and did not express any preference for one technique over the other. This design allowed us to determine how easy it was to master the inferred mode protocol.

To study user preference, we wanted to see whether participants made use of the inferred mode protocol over time. To do this, we designed two interface variants. In the first interface variant, pictured here as Figure 2a, an interface was created that contained four modes: draw, select, delete and smart. The modes were accessed via radio buttons positioned at the top of the screen. The draw mode performed inking in the interface. Select allowed content to be lassoed or clicked on for selection, and translation operations could be performed on selected content for editing. The delete button allowed users to delete entire strokes by drawing a gesture that intersected strokes they wished to delete. Finally, the smart mode button implemented the inferred mode protocol.

When designing the study, one concern we had was that participants using the four-mode interface might never make use of the "Smart Mode" and, therefore, might never see any of the interface techniques that comprise the inferred mode technique. Participants were free to perform the tasks however they wished and we wanted to ensure that at least some of the participants in our study saw the button mediators that invoke computational support. With this in mind, we designed a second interface. The second interface variant (Figure 2b) had only three modes – draw, select and delete. Select and delete functioned identically to select and delete in the first interface. However, the Draw mode was designed to implement the inferred mode protocol, essentially mimicking the behavior of the "Smart" mode in of the first interface.

The task was the entry and editing of a set of simple digital logic circuits. Participants were given an initial digital logic circuit and asked to draw it in the interface. They were then asked to modify the digital logic circuit in specific ways, for example by inserting, deleting, or changing gates. We gave the participants no direction on how to perform the tasks, only what tasks they were to perform in the sketch interface.

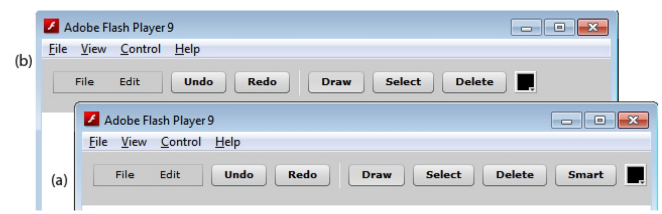

Figure 2: The explicit interface in the foreground (a) and the implicit interface in the background (b)

The study was designed as a between subjects, multi-session observational study. Each participant was assigned to one of the instruction/interface configurations, and remained with that instruction/interface configuration throughout their session. Each drawing and editing session took approximately 45 minutes and completed between three and five sessions until behaviour stabilized.

Eight participants were recruited for this study; none of them had much knowledge of Tablet PC interaction. We used a screen capture application to record participants' actions. The video screen capture allowed us to verify the accuracy of handwritten notes and to quantify the number of times button modes, inferred modes, and mode errors occurred. After the last session, participants were interviewed on their impression of the sketch interface they had used.

Results

Our study revealed information about the protocol in a realistic environment that was not found during the initial evaluation. For instance, learnability is a factor of both previous experience and how the protocol is presented in the interface. Giving the user the option to activate the protocol as a separate mode from the regular features proved to be most effective in incorporating the computational intelligence. In contrast, embedding it directly in the "Draw" mode caused considerable frustration to the users. This frustration is exacerbated by participants' inability to match their gesture input to the output on the canvas. As a result, one of our primary observations is that computational intelligence within interfaces should be viewed as a technique for experienced users in the same way that accelerator keys are a feature geared toward experienced users. Further evaluation of our results can be found in [3].

Conclusion

Many past studies of sketch interfaces give users a very rigid set of tasks for evaluation. Often, there is an assumption that participants would step through a tutorial or learn about the interface in some way before use. However, in many real-world applications, users take a 'walk-up-and-use' approach to software interfaces. In this paper we argue that studies high in ecological validity are needed to reveal important information of new interaction techniques as used in the real world. We present an example study that combines qualitative and quantitative analysis over long periods of time to verify learnability and user preference of Saund et al.'s inferred mode protocol. Results from our study show that properly designed long term experiments can provide important information undiscovered during highly controled, rigid evaluations.

References

[1] K. Hinckley, P. Baudisch, G. Ramos, and F. Guimbretiere. Design and analysis of delimiters for selection-action pen gesture phrases in scriboli. In *Proc of CHI '05*, pages 451–460, 2005.

[2] Y. Li, K. Hinckley, Z. Guan, and J. Landay. Experimental analysis of mode switching techniques in pen-base user interfaces. In *Proc. CHI 2005*, pages 461 – 470, 2005.

[3] M. Negulescu, J. Ruiz, and E. Lank. Exploring usability and learnability of mode inferencing in pen/tablet interfaces. Technical Report CS-2009-29, University of Waterloo.

[4] J. Ruiz, A. Bunt, and E. Lank. A model of non-preferred hand mode switching. In *Proc. of GI '08*, pages 49–56, 2008.

[5] E. Saund and E. Lank. Stylus input and editing without prior selection of mode. In *Proc of UIST '03*, pages 213–216. ACM, 2003.

About the Authors

Matei Negulescu is a graduate student at University of Waterloo’s David R. Cheriton School of Computer Science. He is a member of the Human Computer Interaction lab under the supervision of Edward Lank. Matei’s research interest lies in intelligent sketch interfaces. He is currently researching new interaction techniques for mathematical problem solving systems and performing longitudinal evaluations of new mode switching techniques.

Jaime Ruiz is a graduate student at the University of Waterloo working with Edward Lank. His research interests include mobile interaction and motion kinematics.