How Do Users Adapt to a Faulty System?

Ahmed Sabbir Arif
York University
Toronto, Ontario, Canada
asarif@cse.yorku.ca

Wolfgang Stuerzlinger
York University
Toronto, Ontario, Canada
wolfgang@cse.yorku.ca

CHI'12, May 5–10, 2012, Austin, Texas, USA.
Workshop on Designing and Evaluating Text Entry Methods

Abstract
We investigate how users gradually adapt to a faulty system and how the system error rate influences this adaptation process. We present results of a study that verifies that a user's learning rate to compensate for system errors depends on how erroneous that system is – they learn to avoid erroneous actions faster if errors occur more frequently.

Author Keywords
Text entry; faulty system; user interface; learning.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human factors; measurement; performance.

Introduction
One can observe in many systems that practitioners adapt to a particular system error if it remains in the system for long enough. Once users get accustomed to an erroneous feature of the system, they either actively avoid replicating the sequence of actions that causes the error or start treating it as a feature. We observed this phenomenon in a number of our own text entry related experiments. One theory of learning [2]

Motivation

Various studies established that current recognition techniques, e.g. for gestures and handwriting, are more error prone compared to conventional input techniques. Therefore, most of these techniques provide alternate modalities, especially for error correction. Although, it is generally assumed that users gradually adapt to a faulty system, no work has verified this. Hence, it is not clear how users really react to a faulty system. Do they adapt to the system by learning how to avoid the erroneous actions? Is there a relationship between the system error rates and the rates at which users adapt to those errors? Answers to these questions are vital as these may provide designers with guidance if a system that is initially faulty may reach an acceptable accuracy rate at some point Also, such results could be used to improve new text entry techniques.

explains this behaviour indirectly as it states that it is vital to avoid mistakes to learn the correct way to operate a system. Recently, a number of new text entry methods, such as touchscreens, digital pens, etc., have opened up new avenues to user interface design. Some of these new methods are error prone. Hence, it is essential to acquire a better understand of how users react to erroneous systems to develop more user-friendly input and interaction techniques. Unfortunately, unlike *human* errors, *system* errors and how users adapt to a faulty system are not well examined. Based on our prior experiments we hypothesise that users gradually adapt to a faulty system's system errors. The system error rate also influences the rate of adaptation to a faulty system. That is, users learn avoiding an erroneous action faster if it occurs more frequently. One example is that if inputting the letter *B* with a gesture recognition technique is more erroneous than for other letters, users will either learn to use a more accurate alternate method, if available, or will take extra care, i.e. extra time, while inputting it. We present a user study that verifies our hypothesis.

An Experiment

Apparatus

We used a Bamboo Pen & Touch tablet with a pen for our experiment. The device has an active pen area of 5.8"×3.6". The orientation of the device was switched to accommodate for left- or right-handedness. We used a custom ActionScript 3.0 application developed with the default Bamboo Mini SDK 2.1. The application was displayed on a laptop computer's 15.4" LCD monitor at 1280×800, see Figure 1. The application used the $1 recognizer [3] to process the pen-based gesture input. The software logged all interactions with timestamps in real-time and calculated user performance directly.

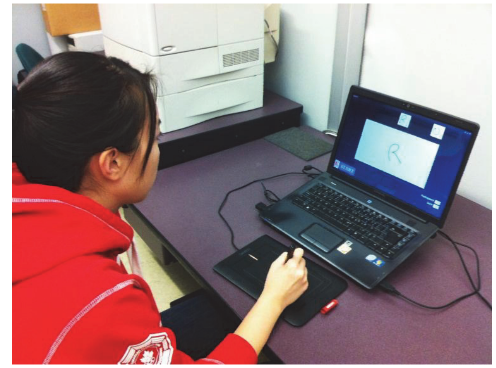

Figure 1. A participant inputting gestures using the Pen & Touch tablet during the experiment.

Participants

12 participants took part in the experiment. Their age ranged from 18 to 27 years, average 22.42. 5 of them were female and 11 of them were right-hand pen users. We only included participants who were not experienced with *Unistrokes* or *Graffiti* in order to eliminate a potential confounding factor.

Procedure and Design

Seven unistroke-based Roman letters: *B*, *D*, *O*, *Q*, *R*, *W*, and *Y*, were used during the experiment. The custom software presented one letter at a time on the computer screen. Participants then had to input the presented letter using a pen on the tablet with either *Graffiti* or *Unistrokes* gestures. The letters were displayed on screen along with the corresponding *Graffiti* and *Unistrokes* gestures. *Graffiti* was presented as the *primary* method of inputting letters, and *Unistrokes* as the *alternative.* That is, participants had to primarily use *Graffiti* to input the letters but were allowed to use *Unistrokes* for letters that were frequently misrecognized by the system. Towards this,

Pilot Studies: Findings

We conducted two pilot studies that helped us to design the final experiment. The pilots used the same gesture-based system and apparatus as the final experiment. We found the following:

Pilot Study 1: Users hardly react to system error rates below 10% per letter.

Pilot Study 2: When writing with an erroneous system, users do use an alternate method to input erroneous letters. Even if users cannot tell error prone letters from non-error prone ones, this still holds. In such a situation, users use the alternate method heavily for almost all letters, just to be safe.

Also, users take extra care while inputting erroneous letters, thus taking more time. Those who successfully identify the most error prone letters usually adopt this strategy.

Graffiti gestures were displayed in bigger size than the *Unistrokes* gestures, see Figure 2. We used *Graffiti* as the primary input method as prior studies [1] showed that novice users favour *Graffiti* over *Unistrokes*.

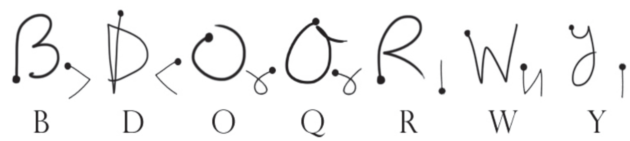

Figure 2. The letters and associated gestures used during the experiment. The larger (primary) gestures are from the *Graffiti* letter set, while the smaller (alternate) ones are from *Unistrokes*. Here, a dot indicates the start of a stroke.

There were two sessions. The *initial* session recorded users' average stroke time, human error rate, and other gesture-related behaviours. It had one block that involved inputting 210 letters. This always proceeded the *final* session to avoid asymmetric skill transfer, especially for faulty letters. The *final* session had three blocks, each one involving entry of 210 letters. Synthetic recognition errors were injected during this session. Three *Graffiti* letters were selected to inject errors at three different rates: 10%, 30%, and 50%, to observe how users input and adapt to these erroneous letters. These three letters were randomly selected for each participant. These three letters were then used during all blocks. The three blocks in the *final* session are designed to analyze how users adapt to a faulty system, as discussed earlier. All letters, including the error-prone ones, appeared the same number of times in the *final* sessions. Error correction was forced during the study. Participants had to keep drawing the letters until it was successfully recognized by the system. In summary, the design was: 12 participants × (*initial* session × 1 block × 280 letters) + (*final* session × 3 blocks × 280 letters) = 13,440 letters. Each participant entered 1120 letters.

Results and Discussion

As the data was not normally distributed, we used a Kruskal-Wallis one-way ANOVA for all analyses.

Attempts

We calculated the average attempts per letter as how many *extra* strokes it took to draw a specific letter. No significance with respect to the number of attempts per letter injected with different system error rates (H_3 = 10.00, p > .05) was identified. This is not unusual considering the mechanism of the experiment, as system errors were systematically injected throughout the blocks to observe user behaviours. Hence, a traditional metric may not be suitable for analyzing accuracy in this case. Further analysis will be conducted regarding this.

The Usage of the Alternate Method

A significant difference with respect to the usage of the alternate method per letter, injected with different error rates, was identified (H_3 = 12.50, p < .0001). Also, the ANOVA found a significant effect for different blocks (H_2 = 210.10, p < .0001). Figure 3 illustrates the mean usage of the alternate input method while inputting letters during each block in the *final* session. Note that participants were instructed to predominantly use the primary method. Figure 3 illustrates that users learned to use the alternate method, when they recognized that the system is not reliable. This validates our hypothesis that users gradually adapt to a faulty system. This is also compliant to our pilot study results that showed that users do a global switch to the alternate method when they cannot identify the erroneous letters.

> **Why These Seven Letters?**
>
> We decided against using short English phrases during the study as using short phrases will require injecting system errors based on a letter frequency table to maintain uniformity. This needlessly complicates the study. Also, typing English phrases with a faulty system causes high level of user frustration, which makes conducting a lengthy and continuous study unrealistic.
>
> We did not use all the Roman letters as it is important that all the letters appear the same number of times during the study to guarantee uniform and comparable adaption rates. A larger number of letters would significantly lengthen a study.
>
> The seven chosen letters were selected based on the most usual methods of drawing them. The intention was to include letters that require relatively similar human effort to draw with *Graffiti* and *Unistrokes*.

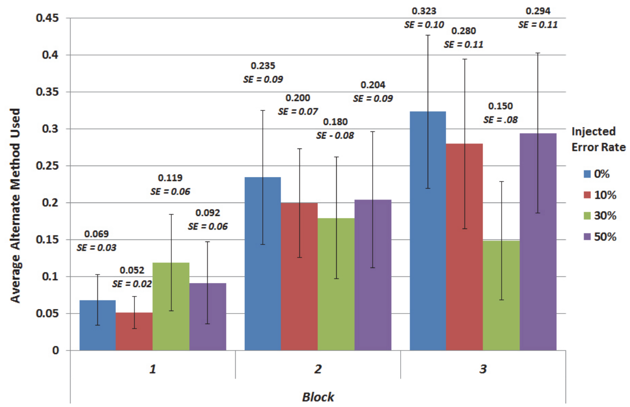

Figure 3. Mean usage of the alternate input method while inputting letters for all investigated system error rates, with standard error, during each block in the final session.

Extra Care while Inputting

We calculated the average time each participants took to draw a particular letter during the *initial* session, which was then used as a baseline during the *final* session. A gesture input was tagged as "extra care" when a user spent more time than the usual (the baseline) to draw that specific letter. We counted the total occurrence of such events for this analysis. The ANOVA identified a significant effect for extra care per letter, injected with different system error rates, (H_3 = 417.97, p < .0001). A significant effect for blocks was also found (H_2 = 6.30, p < .005). Figure 4 illustrates the mean extra care taken by users while inputting letters for all error rates during each block in the *final* session. From Figure 4 it is clear that users learned to take extra care only when inputting error prone letters. The figure shows that the mean of taking extra care decreased for non-erroneous letters (0% injected error rate) and increased for the error prone ones for every block. Also, this increment seems to be proportional to the injected error rate. This validates our second hypothesis that a users' learning rate for a system error depends on that error's occurrence rate.

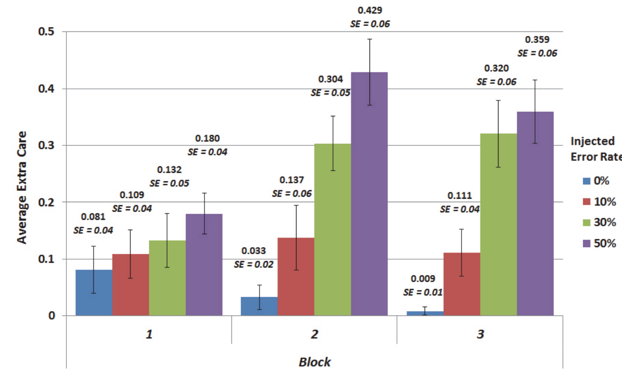

Figure 4. Mean extra care taken by users while inputting letters for all investigated system error rates, with standard error, during each block.

Conclusion

This article confirms that users gradually adapt to system errors. Also, the system error rate influences how users adapt to a faulty system. Users learn to compensate for more frequent system errors faster. Further analysis is of the data is currently in progress.

References

[1] Castellucci, S. J., and MacKenzie, I. S. *Graffiti* vs. Unistrokes: An empirical comparison. In *Proc. CHI 2008*, ACM Press (2008), 305-308.

[2] Newell, A. and Rosenbloom, P. S. Mechanisms of skill acquisition and the law of practice. In J. R. Anderson, ed. *Cognitive Skills and Their Acquisition*, Lawrence Erlbaum, Hillsdale, NJ, USA, 1981.

[3] Wobbrock, J. O., Wilson, A. D., and Li, Y. Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In *Proc. UIST 2007*. ACM Press (2007), 159-168.