Troi Chua

November 18, 2018

Usability Report

I. Introduction

The systems being studied are variations on a time setting application. A user utilizes three navigation keys (left, right, and select) to navigate a menu to set the day of the week and the numerical time. As the interface is navigated, visual and audio feedback, are constantly being provided.

Subjects used two variations of the same time setting application. The systems have slightly different layouts with different parts of the display reacting and providing visual feedback to the user as they enter keystrokes (Fig 1). These visual feedback features are the focus of this experiment as one version may have better stimulus-response compatibility. As the highlighting cells in version 2 are closer to the rest of the interface, subject performance should be better on this version. These differing versions were termed Discrete and Integrated for the placement of the visual feedback.

In order to measure subject performance, two measures were taken. The number of keystrokes was a measure of accuracy and the time it took the subject to complete a trial between the first and last keystroke was a measure of speed.

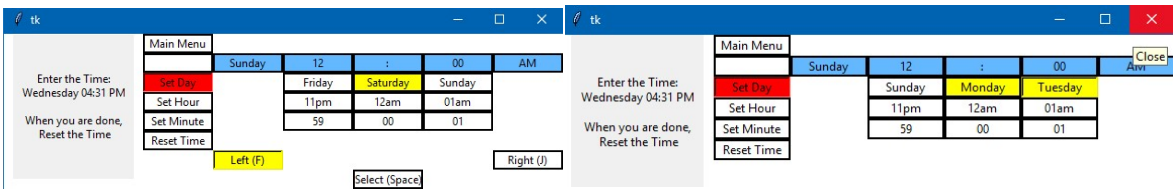

Fig. 1. The clock setting interfaces the subjects interacted with. Discrete (left) had dedicated icons that provided feedback when the subjects pressed keys. Integrated (right) has the cells light up when the key with the same direction is pressed.

II. Methodology

The experiment took place in the Allan Price Science Commons and Research Library (PSC) – a location where science and engineering students study and work. These students were the population that the five subjects were recruited from. Recruitment involved the experimenter going up to students working throughout the library and inquiring whether said individuals could spare about ten minutes (the approximate experiment runtime) and be willing to participate in this study. Those who agreed to participate were all undergraduate males, between the ages of 18 and 22, whom were at least experienced users with computational devices but had zero exposure to the experiment systems. Participants were led to the workstation the experiment was setup on.

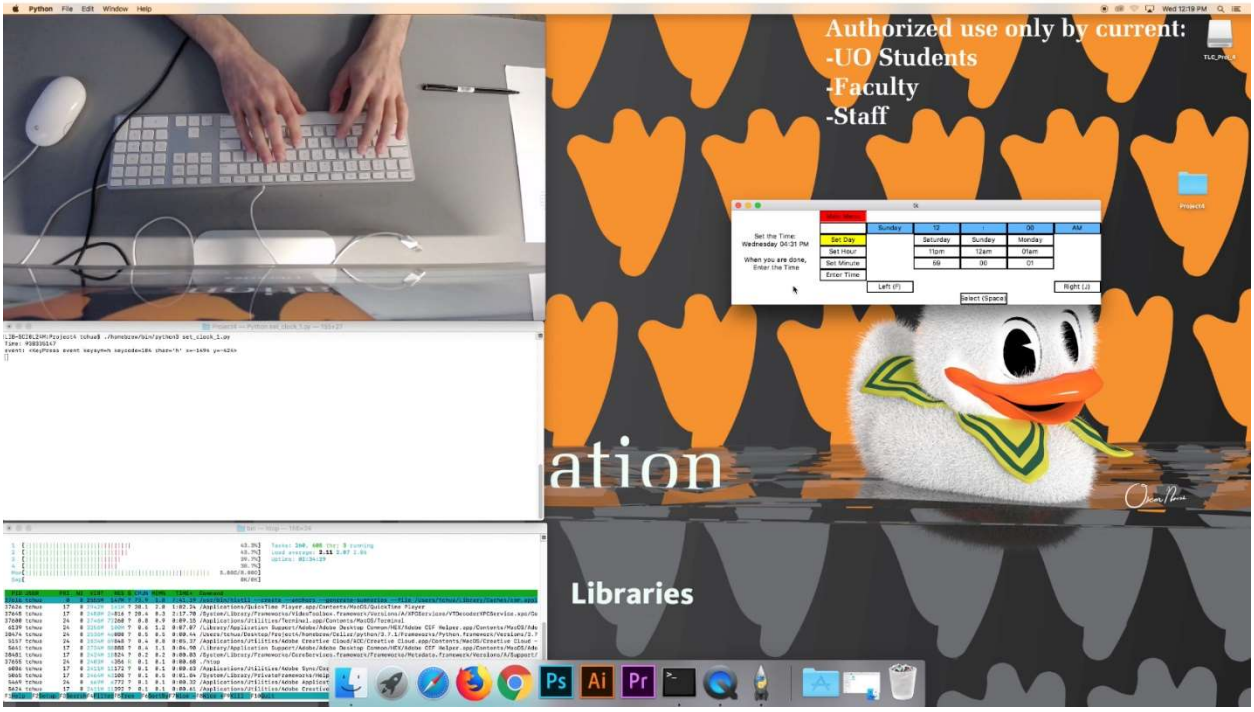

Fig. 2. The display setup for the experiment. Subjects focused on the right side of the display where the time system interface was located. On the left, a QuickTime Player Window records the user's keystrokes while the terminal below does the same. The bottommost terminal uses Htop to monitor system resources during the experiment.

The library grants students access to Windows and Macintosh workstations with necessary productivity software preconfigured on them. The experiment was run using one of the Macintosh systems, and despite having most of the software necessary to run the experiment, the machine lacked the Python dependencies to run the interface. To circumvent this, a portable homebrew program and Python3 distribution on a USB 3.0 flash drive was utilized. An external Logitech c930e Webcam was used in conjunction with QuickTime Player for recording the subjects' physical keystrokes as well as the screen. Subjects were also given a pair of Audio-Technica-m50x's (passive noise isolation) headphones so they could use the audio aspects of the display without disturbing the other library guests.

At the beginning of the experiment, the experimenter read to the participants their rights and were given an informed consent form to study and sign before the experiment could begin.

Once the document was signed, the subjects were then instructed on the procedure they were to follow.

1. The experimenter would explain to the subject the layout and instruct them to focus on the interface, which would be on the right half of the display on the right side of the display.
2. The subject was informed about there task to set the time 10 times (Fig. 3). They were also informed of the navigation keys: J for right, F for left, and spacebar to select.
3. Subjects were informed that they would be doing these 10 tasks twice. The experimenter would start the second version as soon as they were done with the first.
4. Subjects were also instructed to try and set the time with as few keystrokes as possible.

Target Times:

Wednesday, 04, 31, PM

Sunday, 01, 12, AM

Thursday, 06, 42 AM

Friday, 11, 59, PM

Wednesday, 04, 31, PM

Tuesday, 04, 44, AM

Saturday, 07, 18, PM

Monday, 08, 30, AM

Sunday, 00, 00, AM

Wednesday, 04, 31, PM

Fig. 3. The 10 target times that the subjects had to set.

After informing the subject of these tasks, the experimenter would setup the system. After using an online pseudo-number generator on their mobile device to decide which version the subject would begin with, the experimenter would start the corresponding version and adjust the interface window to the right side of the display. The experimenter would then start QuickTime Player's recording functions and instruct the subject to begin. When the subject finished the first version, the experimenter would immediately start them on the second. When the subject completed the second set, the experimenter would shutdown the recording software, ask follow-up questions, and thank the participant.

As stated earlier, there are two versions of the interface: Discrete and Integrated. To avoid one having higher performance due to user practice and experience, the subjects were pseudo randomly assigned so that the first four participants would have half of their members starting with the Discrete version and the other half starting with Integrated version. The last participant was assigned his version order just randomly. The subjects were also not informed that there was a visual layout difference, so their attention would not be influenced into being drawn towards it. There are only 10 trials that are the same across both tasks. All five of the participants received these trials for both versions to ensure consistency for comparing the results and have a consistent baseline to compare to.

III. Results

The data collected for this experiment contains intriguing information about interface design and human strategizing. While not unsurprising, it is interesting to note that perfect accuracy trials are rare in the data. If we start from the beginning, we notice that subjects take their time with the first trial, whether it be starting outright or learning to navigate the interface. This behavior did not change for some subjects when the first trial appeared for their second

round. Speed and accuracy improved swiftly for the subjects as they went along. The subject may have attempted to be accurate as instructed, but they did not always maintain this behavior. There were times, such as during the 5th and 10th trials, which are repeats of the first trial, that the subjects would attempt to enter keystrokes quicker. This resulted in improved speed, but lower accuracy and extra keystrokes from overshooting the correct settings.

The degradation in accuracy performance and improvement in speed performance can be seen in Fig 4. These plots show the temporal progression of participant error. Participants produced more keystroke errors during their second trials, but they reduced their trial time as well.

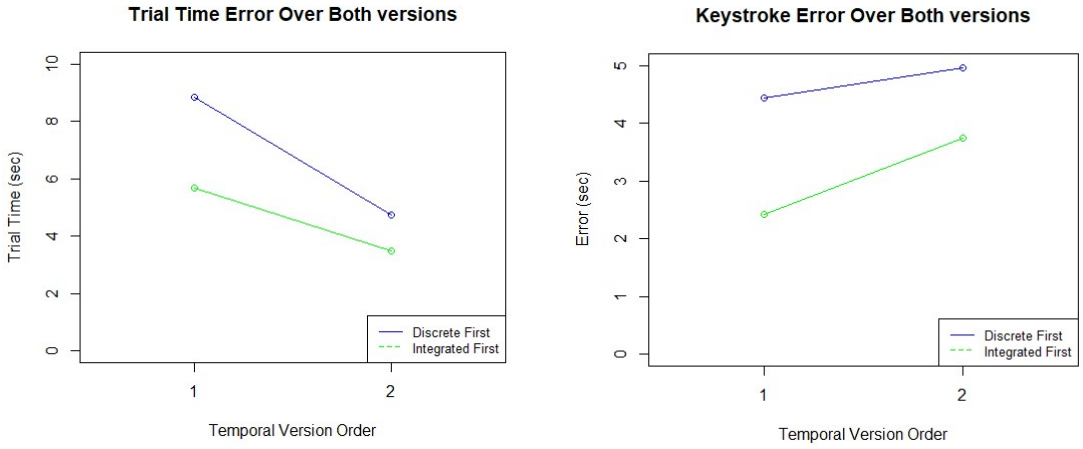

Fig. 4. Participants reduced their time-input error (left) at the cost of making more keystroke errors (right).

Other key behaviors include post mistake optimization. Some of the subjects would make a mistake and go the wrong way for their first keypress for the trial. Depending on the target distance, the subjects would sometimes continue to press the initial wrong keystroke to get to the target from the other direction.

There are a few key bugs that need to be pointed out. There appears to be some difference between how the Graphical User Interface (GUI) framework, tkinter, behaves in Windows and Mac OS. The GUI systems were designed on a Windows 10 system and run on a Mac OS X 13 Macintosh. In Windows, certain functions, such as holding the key for continuous keystrokes, are not functional on Mac OS. This may affect strategies and usability because holding the button may seem more intuitive to use for longer strings than rapid-fire keypresses. Some subjects also reported audio only being functional on one of the systems. The system that was reported to have functioning audio feedback was not consistent, indicating a possible runtime malfunction or OS compatibilities.

A more impactful error has to do with trial 9. This trial has a bug in the prompt. It is supposed to request the user enter Sunday 12:00 AM, which is the value the clock sets to after each trial. This time only takes 2 keystrokes, which also results in swift reaction times. However, the clock sets itself to Sunday 00:00 PM, which is not a time in the standard language of

contemporary timekeeping. Many subjects were confused and paused to ponder this prompt as well as explore the hour interface, costing them both speed and accuracy. A few students immediately went to 12:00 am or 12:00 pm, but some took time to ponder this mistake.

The abnormality of the 9th appears to have had a great effect on the participants such that it had negatively impacted their performance for the final trial. Fig. 5 shows the participants' keystroke error and the performance for the final two trial can be seen. It was decided that the last two trials would be excluded from the overall analysis due to the confounding nature the 9th trial had.

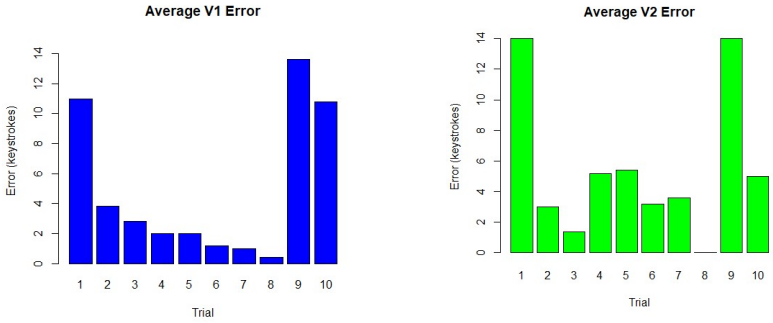

Fig. 5. The average keystroke error including the last 2 trials. The Discrete system (left) shows a highly impactful effect trial 9 had on subsequent performance

IV. Discussion

The data collected from the subjects conducting these trials was analyzed using the R-Statistical Computing Language. To simplify things, an average of each measure was taken from all subjects for each trial. The experimental design instructed the participants to conduct tasks which varying keystroke minimums and times needed to accomplish them. Due to this factor, the data that was analyzed focuses more so on the keystroke error count and time over the baseline time it took the experimenter accomplish the tasks.

The beginning case is the least interesting because a more expected outcome. They were relatively inaccurate and a quite a bit slower than the other trials. This tutorial trial was enough for most of the subject to get in the groove and understand how to work the interface as shown by the increasing performance levels.

Fig. 6. shows the keystroke error improvement as the trials go. The Discrete system shows a more stable improvement curve as well as overall better performance. These performance curves are similarly seen in the temporal error plots. The progression here is not as smooth and the difference in times is negligible.

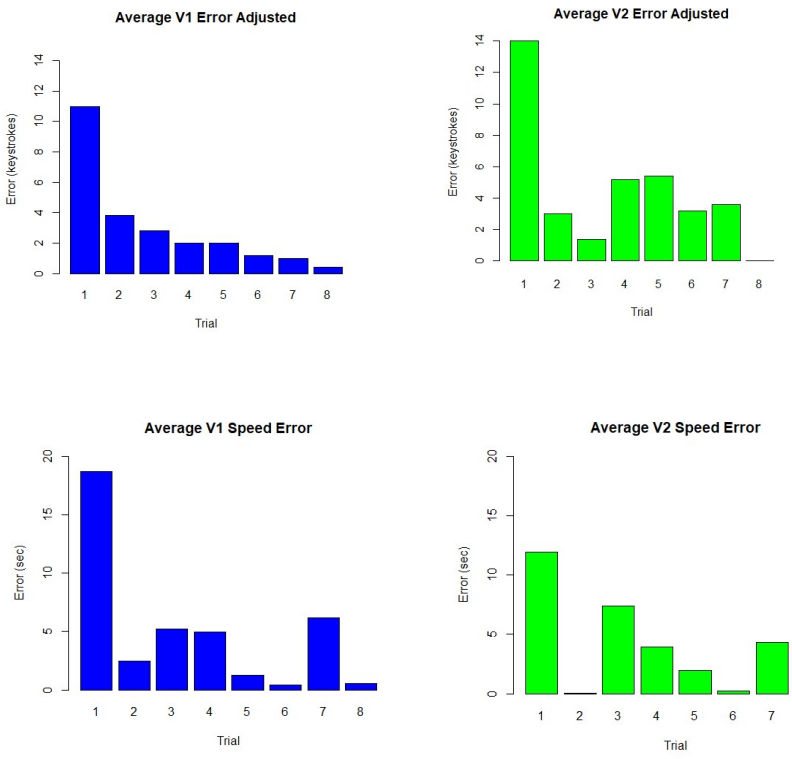

Fig 6. The errors of both systems over the 10 trials. The Discrete system are the left plots and the Integrated system plots are on the right. The Discrete system shows smoother keystroke error reduction (top plots) while the speed errors are relatively similar.

The R-statistical software was used to calculate the mean values for the study. The Discrete System had better accuracy performance with participants taking less overall keystrokes and having and key error, but the integrated system had better speed performance as participants took less time to accomplish the tasks. These number are displayed in Fig. 7. It is apparent that these differences are statistically insignificant. The data suggests that there is no significant difference between the performance that can be achieved with these systems. However, no true conclusions can be drawn: the sample size is too small, so not enough data was collected to make determine a significant difference between the interfaces.

Discrete Averages	**Integrated Averages**
• Keystrokes = 38.675 strokes • Trial Time = 20.933 sec • Key Error = 3.025 strokes • Time Error = 4.969 sec	• Keystrokes = 40.225 strokes • Trial Time = 19.637 sec • Key Error = 4.475 strokes • Time Error = 3.809 sec

Fig. 7. The averages of keystrokes, trial times, and their errors for both systems.

V. Conclusion

This experiment brought to light some interesting behaviors and strategies that users employ when undergoing a repetitive task. Even if the user is consciously aware and determined to go for accuracy, they may try to speed up to increase productivity, thus increasing the margin and probability of error. A key thing to note is that the hypothesis focused on the difference in visual layout and feedback stimuli. When the subjects were inquired about their thoughts on the system and the experience, they noted many things about the choice of buttons or the overall system, or that some of the audio feedback was redundant. When they were inquired about noticing the difference between the interfaces, almost all of them spoke of there being an audio difference, which there was not one intentionally programmed in. Not one of the subjects noted the difference in visual layout which could possibly indicate that these features may possibly hold little significance while conducting the task. This indicates that there may be some hierarchical value to the conjunction of having both sound and visual stimuli. This also indicates that the features of visual feedback provided in the interfaces need to be tweaked so they have a more significant impact. To explore this further, the experiment needs to be debugged then expanded so that more data is collected from more diverse subjects and tasks.