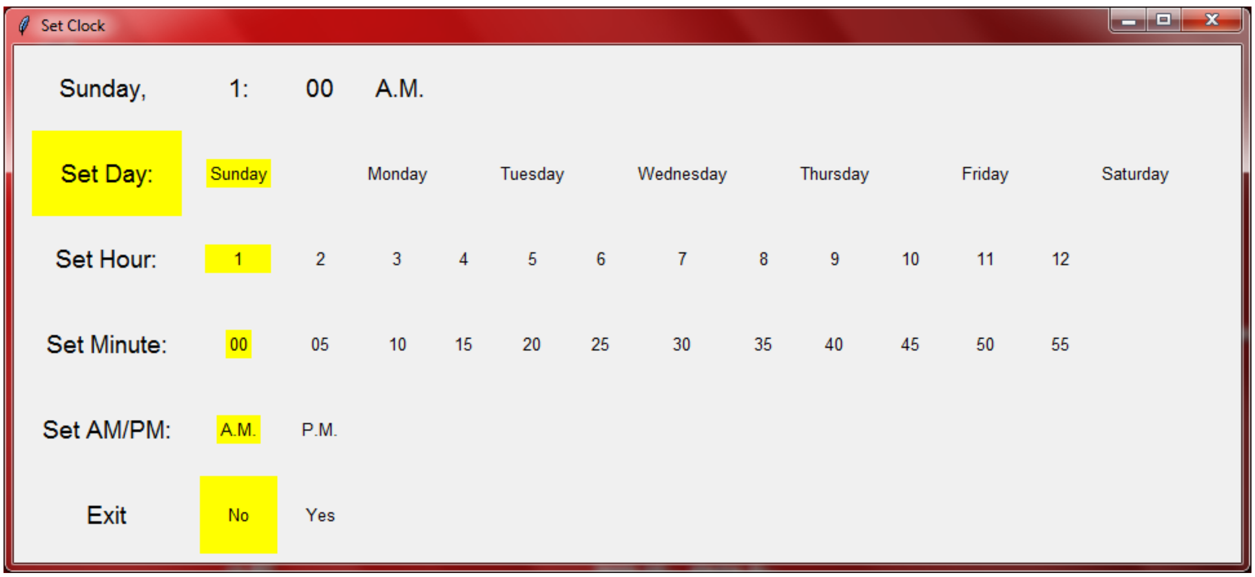
Author: Cole Vikupitz
Last Modified: 11/27/2017

User Observation Study - Usability Report

1. Introduction

The system that was used in this user observation study was an application where users manually set the day and time. The application is a GUI (graphical user interface), and the user uses the keys J, K, and the space bar to cycle through the application's modes, and also to select each individual hour, minute, day, etc. The current mode and the user's selections are shown by highlighting the selected text(s) in bright yellow. The application was written in Python and uses the TkInter library for the interface. The application described is shown below.



(Above) The application the users used in the observation study.

The goal of this study is to observe user's interactions with a particular system and observe how quickly and accurately the user can accomplish a given task. In particular, the goal is to observe how different key mappings will affect the user's accuracy and speed in using the system. Two systems have been built, but both have different key mappings, therefore a different level of stimulus-response compatibility. These are outlined as such:

System #1: J (Backward), K (Forward), Space (Select)

System #2: J (Backward), L (Forward), K (Select)

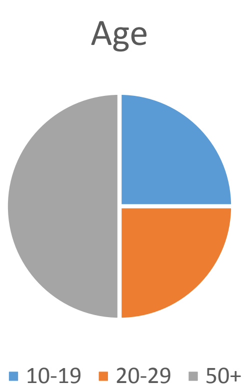
I hypothesize that the participants will perform the tasks faster and more accurately with system 1 than 2. This is because I believe that there is a better affordance to selecting or confirming something with the space bar, rather than one of the letter keys. Users may also find system 2 harder to use having the select button in-between the forward and backward buttons, slowing them down and reducing their accuracy.

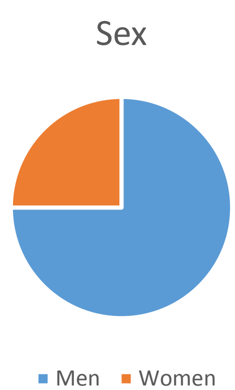
To measure the results, I will be testing the difference in time the users take to accomplish the set of tasks for both systems I will also be measuring the difference in the number of keystrokes the users take in order to accomplish the tasks. I hypothesize that users will accomplish the tasks for system 1 30-60 seconds faster than for system 2. I also hypothesize that users will press more keys on system 2 than system 1 due to the lower level of stimulus-response compatibility.

2. Methodology

2.1 Participants

Four people participated in this observation study; I could not find a fifth person willing to participate or be recorded in time. Two of the participants are university students, one a biology major, the other a physics major. The other two are older adults; one is a business owner, and the other is a retailer at a sports shop. All four participants have basic knowledge on how to use a computer. More demographic information given below.





2.2 Setting

One user participated in the study in an apartment unit at a table. The other three users participated in a home office setting at a desk. The users were kept isolated from anyone else and any other distractions (a TV, other people talking, etc.) while the study was taking place.

2.3 Materials

• *Laptop computer* for the users to test the systems on.

• An *IPhone* to record the user's interaction with the systems.

• Two *bash scripts* that run the systems and record all the user's keystrokes into log files.

• An *instruction sheet* for users to follow and execute the tasks on the system.

• A *notepad* to record usage patterns, and user's answers to post-experiment questions.

2.4 Experimental Design

The two conditions being tested, as stated earlier, are the different key mappings for the two systems. Therefore, the variable being manipulated (the independent variable) is the key mapping. The results that occur due to these changes are the user's speed and accuracy in carrying out the tasks (the dependent variable). To test these conditions, users will use two systems that are the same in appearance and functionality, but differ in the key mappings that are used.

This study used a within-subjects design, as in, the user was exposed to all levels of the independent variable manipulated. This was done to get as many results for both systems as possible. Doing a between subjects design would require more time and volunteers to get good results. Since each participant was exposed to both systems, it was important to ensure that the order in which the systems were introduced did not stay the same. Therefore, the order was alternated between each user (first user used system 1 then 2, user 2 used system2 then 1, etc.).

Key Mapping

System 1	System 2

Key Mapping

System 2	System 1

(Left) The experimental design used. Each user was exposed to both system, each with a different key mapping. The order of the systems used is shown as such. The top design shows that half of the users used system 1, then system 2. The other half used the bottom design, they used system 2 first, then system 1.

For each system the user interacted with, they were given a set of 7 tasks to accomplish organized into a trial. The results recorded show how long it took the users to complete the given trial. Since the user interacted with two systems, they were given two trials, both similar in the tasks the user must accomplish. The user was to set three specified times, and have each displayed to the user, then the user was to exit the system. All the times the users were given to set were randomized with a random number generator.

2.5 Procedure

The observation study was carried out in the following steps:

1.) Assert that the participant meets all the requirements (not a CIS student, nor enrolled in CIS 443/543).

2.) Introduce myself, give a brief summary of the observation study and what the user will be doing.

3.) Give user consent form, allow user to sign. Inform they may leave at any time.

4.) Give the user the first trial on the selected system, start the system on computer. Record and allow the user to perform the tasks. Intervene or provide assistance when needed.

5.) Repeat step 4, but replace first trial with second trial.

6.) Ask the participant the debriefing questions, record the answers.

7.) Thank the user for participating. Any other question(s) user may have may be asked and answered.

2.6 Threats to External Validity

• *All the participants are in the same age group*. This can cause the results to reflect poorly on how the general population learns or uses the system. This is eliminated by recruiting people of diverse ages.

• *All participants use the same operating system*. This observation study was done on a Windows machine, and Windows users may learn the system faster than a user who uses a different operating system. This is eliminated by recruiting people who use different systems.

• *Every study is conducted with both systems introduced in the same order (e.g. every participant used system 1 before 2).* Users may find system 2 easier to use after just using system 1, which introduces learning bias. This is eliminated by randomizing the order of trials (system 1 introduced before system 2, and vice versa).

3. Results

Time to Complete Trial	**System 1**	**System 2**
User 1	148 sec.	132 sec.
User 2	78 sec.	127 sec.
User 3	235 sec.	N/A[1]
User 4	125 sec.	270 sec.
Average Time	146.5 sec.	176.33 sec.

Number of Keystrokes to Complete Trial	**System 1**	**System 2**
User 1	81	82
User 2	67	113
User 3	103	N/A
User 4	77	125
Average Number of Keystrokes	82	106.67

[1] User became frustrated with the test system, and ended the experiment. Results were therefore thrown out.

Accuracy[2]	**System 1**	**System 2**
User 1	48.1%	50.0%
User 2	83.6%	48.7%
User 3	60.2%	N/A
User 4	72.7%	38.4%
Average Accuracy	66.2%	45.7%

4. Analysis

After each study was complete, each participant was asked a set of debriefing questions about the study. When asked if the system always responded as expected, 2 said yes, and the other 2 said no for system 2 but yes for system 1. When asked if one system was easier/harder to use, 3 participants said system 1 was easier to use, while one participant said both were easy, but required learning first. Finally, when asked which system they preferred, 3 participants said they preferred system 1, while one participant liked system 2 due to all the keys being next to each other.

Based on these answers given, and the data collected from the trial runs, it appears that system 1 has better usability than system 2. The users, on average, performed the tasks faster and more efficiently with system 1 than 2. One of the participants even decided to quit the observation study due to frustration with system 2. Due to these results, it would appear that the key mapping for system 1 is superior to system 2's, in that it has a higher stimulus response compatibility. One user in particular remarked that system 1's key mapping made more sense than system 2's when asked if the system always responded as expected. This would support the idea that system 1's key mapping has a better affordance than system 2's.

In general, system 1 is more usable than system 2, and seems to meet its usability goals. System 1 did not hinder the user's ability to complete the tasks at hand. The users did need to experiment with the keystrokes a few times in order to learn how the system worked. However, users spent less time learning the keystrokes and making mistakes on system 1 than they did on system 2, and therefore were able to complete the tasks quicker and more accurately.

To improve on the interface, the menu feature should be implemented into the system; users should have the ability to select the mode they're in rather than keep pressing the select button. The interface should also highlight the complete time when the user updates the time, to let them know the mode they are in. During observation, many participants were uncertain what to press after updating the time.

[2] Calculated by dividing the minimum keystrokes required to complete the trial by user's total keystrokes.

5. Conclusion

In conclusion, system 1 was more usable than system 2 in the sense that users were able to accomplish tasks faster and more accurately. This suggests that having a strong stimulus-response compatibility is a crucial element in building an interface, and to the average user's experience. Having poor SR compatibility, on the other hand, will leave users unsatisfied, and often frustrated, as I have experienced during the study.

One thing that I have learned from conducting the observation study was that users do not like to learn interfaces very much. The less time users spend leaning and making mistakes, the more satisfied they are with the system. With ease of learning and usage, users will be quicker and more accurate when using the system.