Author: Mamtaj Akter
Last Modified: December 1, 2019

User Observation Study – Usability Report

1. Introduction

This user observation study analyzes two interactive systems where users play a chapter content of an audiobook using. Both the systems have pushbutton input and auditory output interface. The user uses the keys <SPACE>, <J>, <K>, <L> and <;> for the inputs. The goal of the systems is to play Chapter Two content of an audiobook. The functionalities of system 1 and system 2 are described below.

System 1: *<J> (Scroll Backward), <K> (Scroll Forward), <Space> (Select), <L> (Help),<;> (Quit)*
Mode Order: "Select Book" <=> "Select Chapter" <=> "Continue Reading" or "Content Not Available"

System 2: *<J> (Scroll or Control), <K> (Content Availability), <Space> (Select), <L> (Help), <;> (Quit)*
Mode Order: (*"Select Book" <=> "Help")* <=> (*"Select Chapter" <=> "Help")* <=> (*"Continue Reading" <=> "Help")*

System 1 and System 2 has the same goal – select a book titled "The Design of Everyday Things", and then select Chapter Two: "The Psychology of Everyday Things" to play its content. In both systems, after scrolling through the book titles and chapter names, <Space> is used to select the book or to select the chapter. <L> key is used to play help messages and the function of <;> key is to exit from the current mode/menu and to go back to the previous mode/menu. For example, when a user is in "Select Chapter" mode, <;> keypress leads the user to the "Select Book" mode.

There are two major differences between these two systems. In System 1, users can scroll through the book titles and chapter names using <J> and <K> buttons to scroll backward and forward respectively, while System 2 has only one button <J> to scroll and <K> does not have any function. System 1 has so many useful instructions and feedbacks in every step compared to System 2 where very minimal auditory instructions were given. For example, as soon as users run the application, system 1 plays all the buttons functionalities including the command "Select book". However, System 2 only plays "Select Book, Press L for help".

Both of these systems have some shortcomings. The system 1 lets the users scroll through all the chapter names of the audiobook even though chapters' contents are not available. When the user selects a chapter other than chapter 2, the system enters into the "Content Not Available" mode and therefore, it requires users to exit from that mode to select another chapter. This induces more keystrokes and thus more time. On the other hand, system 2 gives the first audio instruction "Select book, press L for help" and when user presses <L>, the system enters into a help mode and the functions of the keys get changed: <J> becomes a key to learn about the controls that are available in the system, <K> key gives the auditory message on what book and chapter is available in the system, <L> key plays the help message and <Space> does not play any auditory message. Here, there is no instruction given what mode the user is currently in and how to exit from the help mode, which can cause users to press more keys and spend more time.

Since I predict that with system 2, users may get stuck in help mode and might get frustrated and there is a good chance of giving up in the middle of the study, I added an additional auditory message in System 2. When user enters into the "help mode", user may not have any idea in which mode he/she is currently in and thus might try pressing <Space> to select, I added an auditory message: "Press <;> and then press <Space> to select" for the functionality of the <Space> key in Help mode. I named this second version of System 2 as System 3.

I hypothesize that the participants will perform the tasks faster and more accurately with system 1 than 2. This is because I believe that its easier and faster in scrolling backward or forward with the <J> and <K> key presses, rather than using only one key <J> to scroll of system 2. Users may also find system 2 harder to use for having the extra but unhelpful functionality of entering into "Help Mode". I also hypothesize that using System 3 will be faster than system 1 because users will get clear instruction on how to select a book or chapter instantly when they enter in the help mode.

The goal of this study is to observe the users' interactions with two particular systems and observe how quickly and accurately the user can accomplish a given task with those systems. In particular, the goal is to observe how incorporating extra functionality/mode will affect the user's accuracy and speed in using the system. Three systems have been built, but all have different mode settings, therefore a different level of challenge in using and learning the systems.

To measure the results, I will be testing the difference in the 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. Then the practice effects will be measured, such as how practicing once with the systems helps a user to perform in a subsequent block with the same system and different system. I hypothesize that users will accomplish the tasks for system 1 40-80 seconds faster than for system 2. I also hypothesize that users will require more keystrokes on system 2 than system 1 due to the option to enter into the help mode anytime the user presses <L>

for help. However, system 3 will require less time and keystrokes than system 1 and 2 both. I hypothesize that system 1 and 3 has more affordance than system 2 and thus users will require fewer keystrokes to accomplish the task with system 1 and 3. Once the user gets exposed to at least one system, I hypothesize that users can accomplish their tasks with less effort with System 1 because it is easier to use a system once users learn about the system goal.

2. Methodologies

2.1 Participants

Twelve people participated in this observation study. All twelve participants have basic knowledge on how to use a computer. Figure 1 depicts some demographic information of the participants.

User 1: Age: 65, Male, Retired Missionary, Not Experienced with Audio Book, Tested System 1 & 2.
User 2: Age: 25, Female, Masters Student-International Relations, Not Experienced with Audio Book, Tested System 1 & 2.
User 3: Age: 46, Male, Ph.D. Candidate-Political Science, Experienced with Audio Book, Tested System 1 & 2.

User 4: Age: 26, Male, Ph.D. Student-Economics, Not Experienced with Audio Book, Tested System 2 & 1.
User 5: Age: 27, Male, Ph.D. Candidate-Economics, Experienced with Audio Book, Tested System 2 & 1.
User 6: Age: 43, Male, College Student-Freshman, Not Experienced with Audio Book, Tested System 2 & 1.

User 7: Age: 29, Female, Ph.D. in Media Studies, Experienced with Audio Book, Tested System 1 & 3.
User 8: Age: 30, Female, Masters in Journalism, Experienced with Audio Book, Tested System 1 & 3.
User 9: Age: 31, Male, Ph.D. Candidate-Cinema Studies, Experienced with Audio Book, Tested System 1 & 3.

User 10: Age: 39, Female, Ph.D. Student-Political Science, Not Experienced with Audio Book, Tested System 3 & 1.

User 11: Age: 41, Male, Ph.D. Candidate-Linguistics, Experienced with Audio Book, Tested System 3 & 1.
User 12: Age: 13, Female, Middle School Student, Experienced with Audio Book, Tested System 3 & 1.

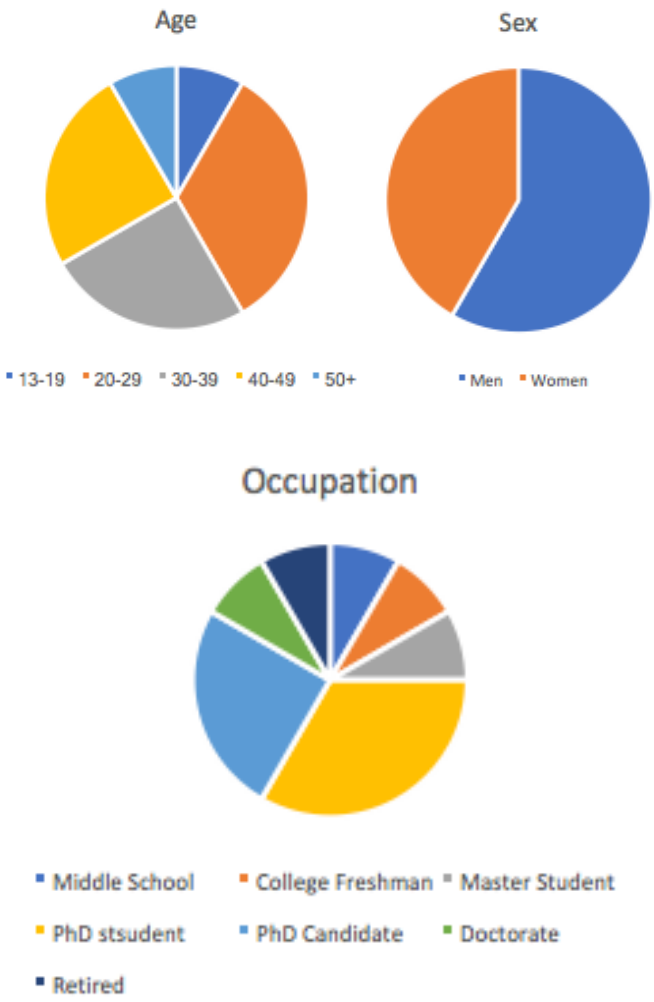

Figure 1: Participants Age, Sex, and Education Level

2.2 Setting

The study took place in variety of locations and time of the day. The users were kept isolated from anyone else and any other distractions (a TV, mobile phones, other people talking, etc.) while the study was taking place.

2.3 Materials Used

1. A MacBook pro for the users to test the systems on.
2. An iPhone to record the user's interaction with the systems.
3. A notepad to record usage patterns, and user's answers to post-experiment questions.
4. An instruction sheet for users to follow and execute the tasks on the system.
5. A copy of the informed Consent Form.
6. Terminal to run the applications.

2.4 Experimental Design

The independent variable that is being manipulated in the study of system 1 and system 2 is the presence of "Presence of Enough Auditory Feedback" in system 1. However, in the study of system 1 and System 3, the independent variable that is being considered is "Simplicity and Straightforwardness in Design" in system 3. Not having helpful auditory instructions or having an extra mode "Content Not Available" may affect the user's speed and accuracy in carrying out the tasks. And therefore, the speed, number of keystrokes and accuracy are the dependent variables. To test these conditions, users will use three systems that are the same in terms of the goal – Select Chapter Two of an audiobook to play its content. But these three systems differ in functionality and usability. System 1 has an additional "Content Not Available" mode, whereas System 2 and System 3 have a "Help mode", without guidance on how to exit from that mode, other than the general modes "Select Book", "Select Chapter", and "Continue Reading". However, System 3 is an advanced version of System 2 with an auditory message "Press <;> and then press <Space> to select" when a user enters into the help mode and presses <Space> to select.

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 all the systems as possible. Doing a between-subjects design would require more participants 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 participant group used system 1 then 2, user 2 used system 2 then 1, etc.) as shown in Table 1. Each participant group has 3 users.

Participant Group	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6
1	System 1	System 2	System1	System 2	System 1	System 2
2	System 2	System 1	System 2	System 1	System 2	System 1
3	System 1	System 3	System 1	System 3	System 1	System 3
4	System 3	System 1	System 3	System 1	System 3	System 1

Table 1: The orders of the Systems that are followed for each participant group

For each system the user interacted with, they were given a set of six tasks to accomplish organized into a trial. The experiment also recorded the time how long it took the users to complete the given trial. Since each user interacted with two systems, they were given two trials, all are similar in the tasks the user must accomplish. Then users will repeat Trial1 and Trial 2 for two more times. Thus, in total, there are six trials were performed by each user. Each Trial has the following set of six tasks:

1. scroll through the available book titles
2. select the book titled "The Design of Everyday Things"
3. scroll through the chapter names of that book
4. select chapter two "The Psychology of Everyday Things"
5. play the content of chapter two
6. finally, exit from the program

2.5 Procedure

The observation study was carried out in the following steps:

1. Introduce myself.
2. Assert that the participant meets all the requirements (not a CIS student, nor enrolled in CIS 443/543).
3. Give the user consent form, allow the user to sign. Inform they may leave at any time.
4. Give a brief summary of the observation study and what the user will be doing.
5. Tell the user that the study will take around 20 minutes and they can quit at any time.
6. Ask request to think aloud as he/she does the task.
7. Give the user the first trial on the selected system, start the system on the computer. Record and allow the user to perform the tasks. Intervene or provide assistance when needed.

. 8. Repeat step 7, but replace the first trial with the second trial.

9. Ask the participant the following questions and record the answers:
    a. What did you think of the task?
    b. Did the systems always respond to your keypresses as expected?
    c. Was either system particularly easy/hard to use?
    d. Which system did you prefer?
    e. Can you give a point 1-10 to each of the systems as per your satisfaction?

10. Thank the user for participating.
11. Any other question(s) the 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 the participants are in the same gender group.* This can cause the results to reflect poorly on how the general population of different gender learns or uses the system. This is eliminated by recruiting an almost equal number of men and women.

• *All the participants are Graduate Students.* Graduate Students may learn the system faster than the users who are not students. This is eliminated by recruiting people from diverse occupations. I have recruited people from middle school to college freshman and from masters/Ph.D. student to Doctorate and retired person.

• *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).

• *All participants used the system at the same time of the day and location.* People may learn the system quickly in their own residence/offices because they may find the study environment in their favor. People may also learn the system faster and make fewer errors when they are either done with the school work or office hours. This is eliminated by conducting this observation study at two places: my office at Deschutes Hall, University of Oregon and users' apartment units. Five users used the system in the afternoon hours and two users participated during the morning hours and the rest of the participants used the systems in my office and time was after 6 pm.

3. Results

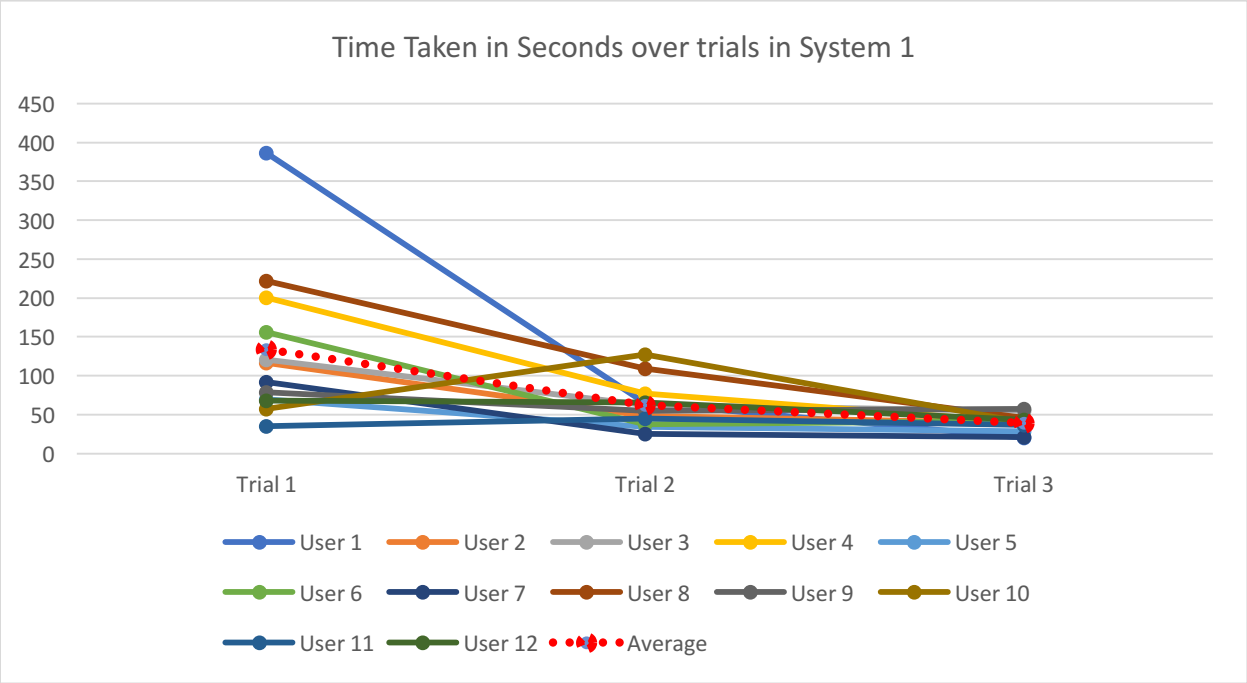

Figure 1.1: Time taken in Seconds in System 1 over the number of Trials

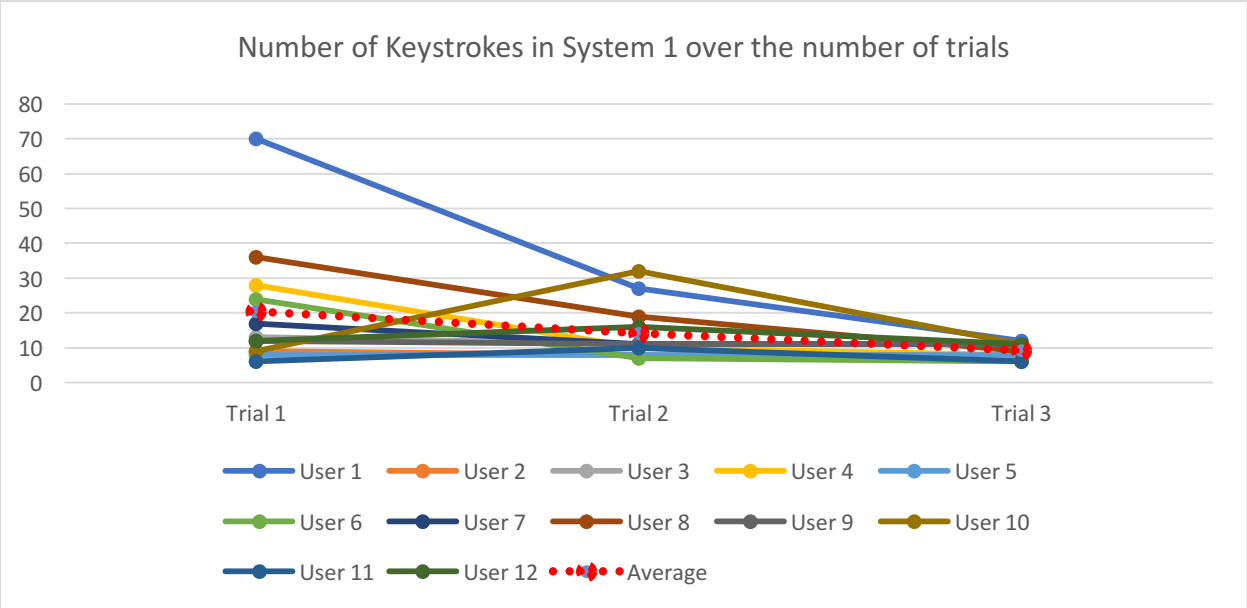

Figure 1.2: The number of Keystrokes in System 1 over the number of Trials

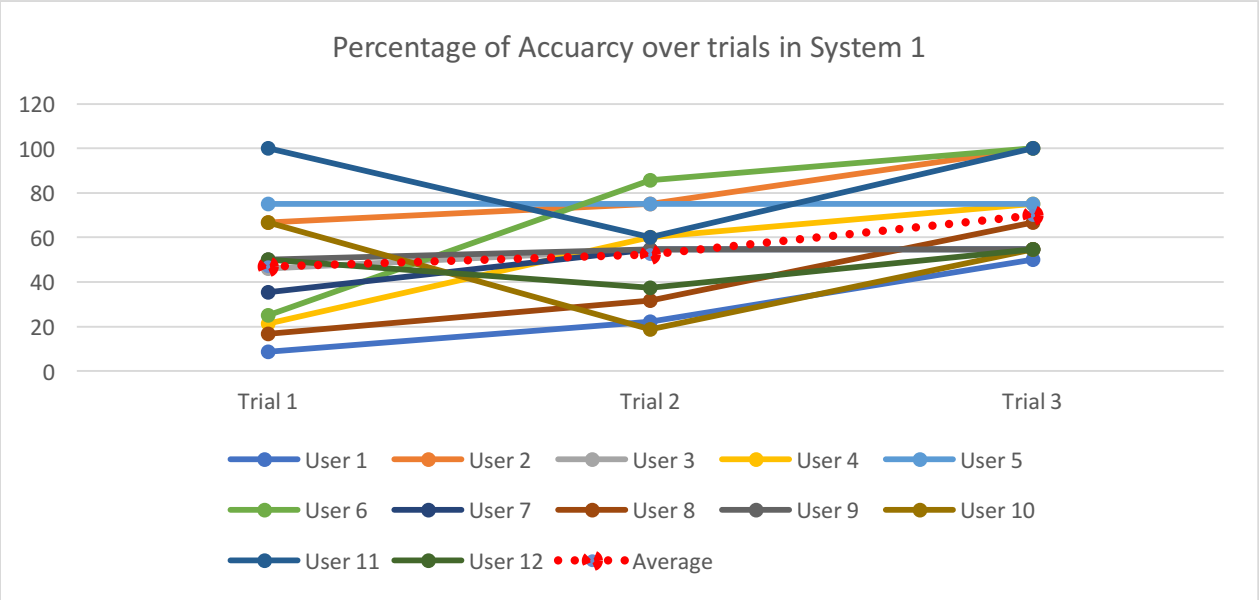

Figure 1.3: The percentage of Accuracy in System 1 over the number of Trials

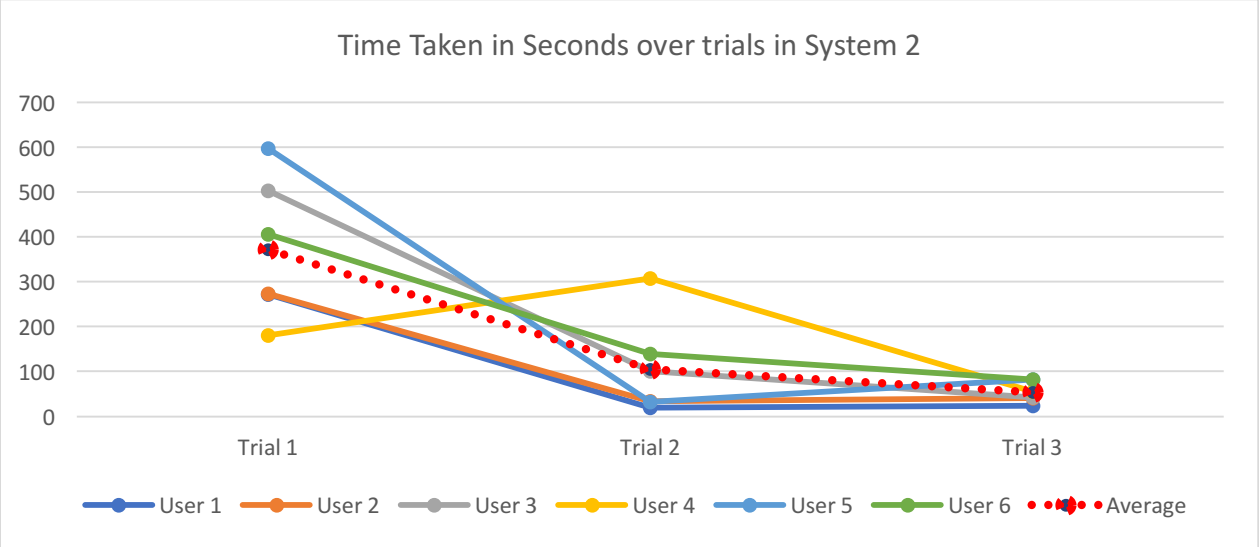

Figure 2.1: Time taken in Seconds in System 2 over the number of Trials

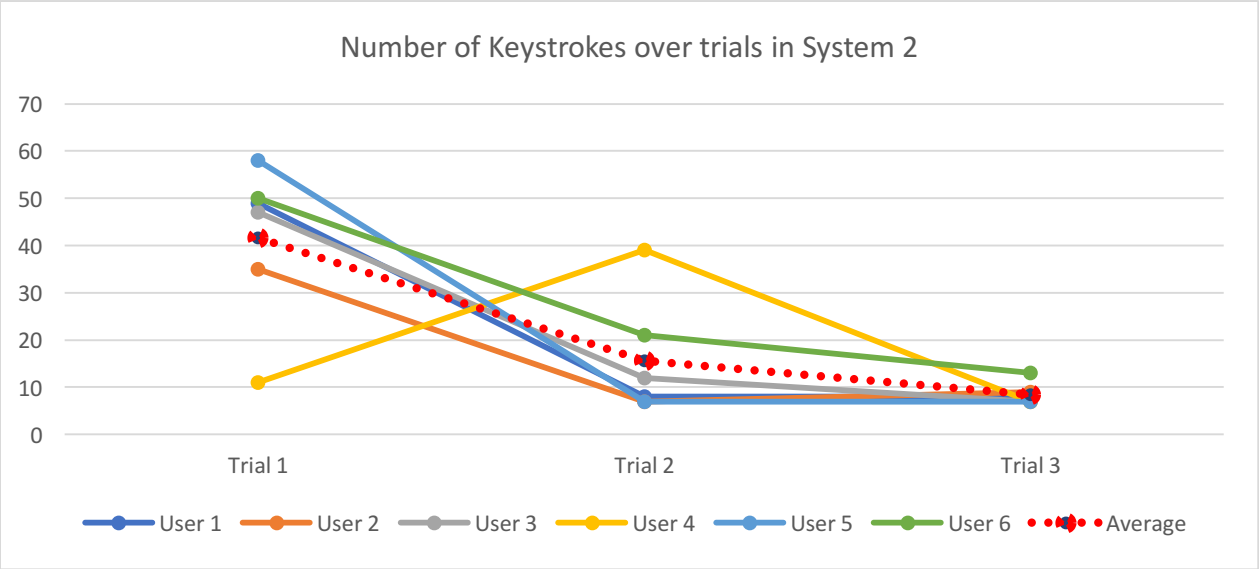

Figure 2.2: The number of Keystrokes in System 2 over the number of Trials

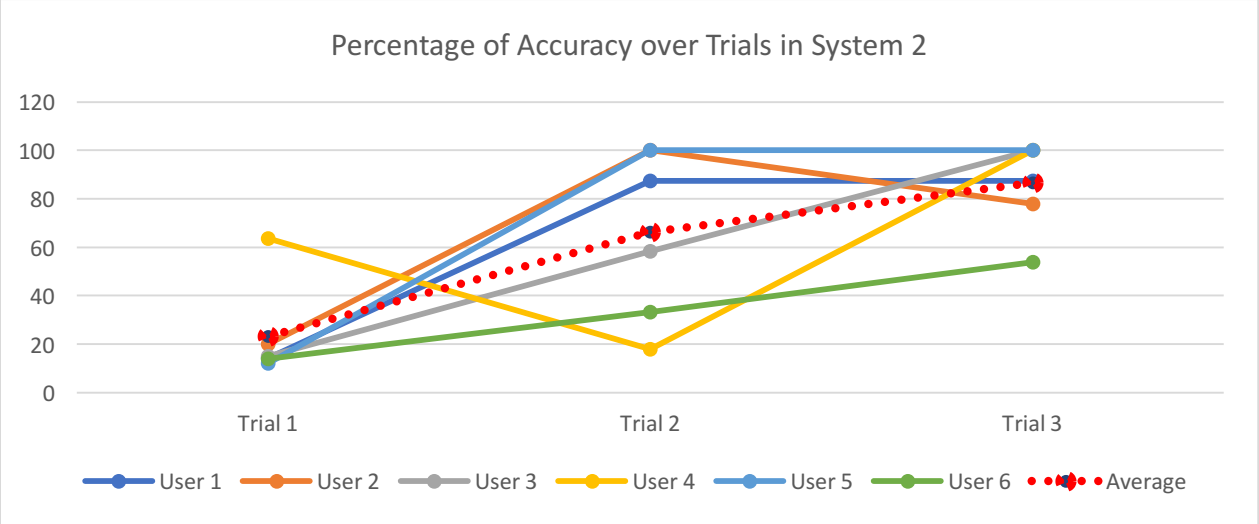

Figure 2.3: The percentage of Accuracy in System 2 over the number of Trials

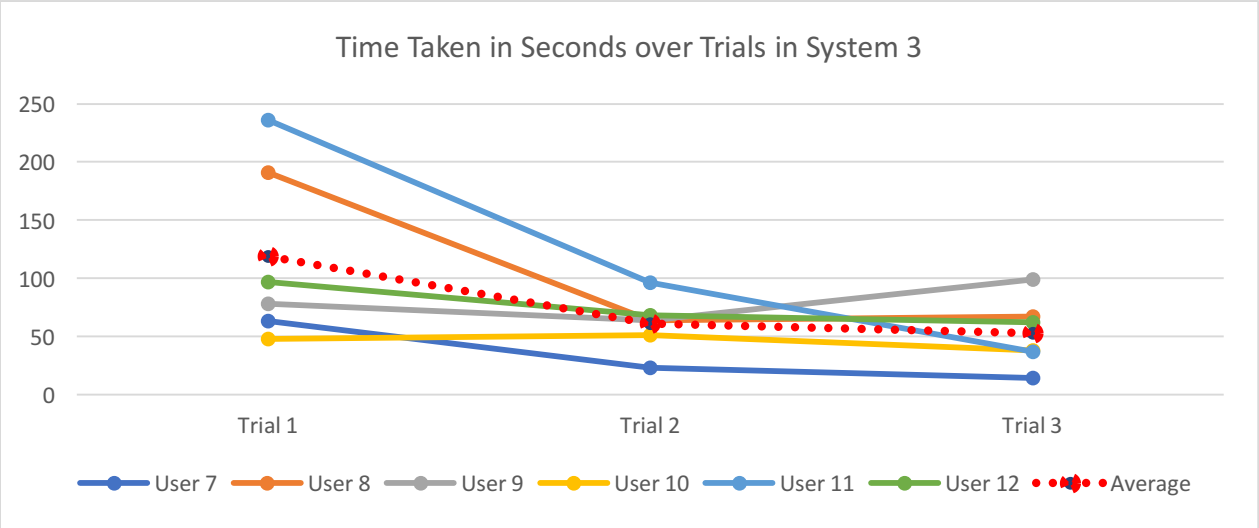

Figure 3.1: Time taken in Seconds in System 3 over the number of Trials

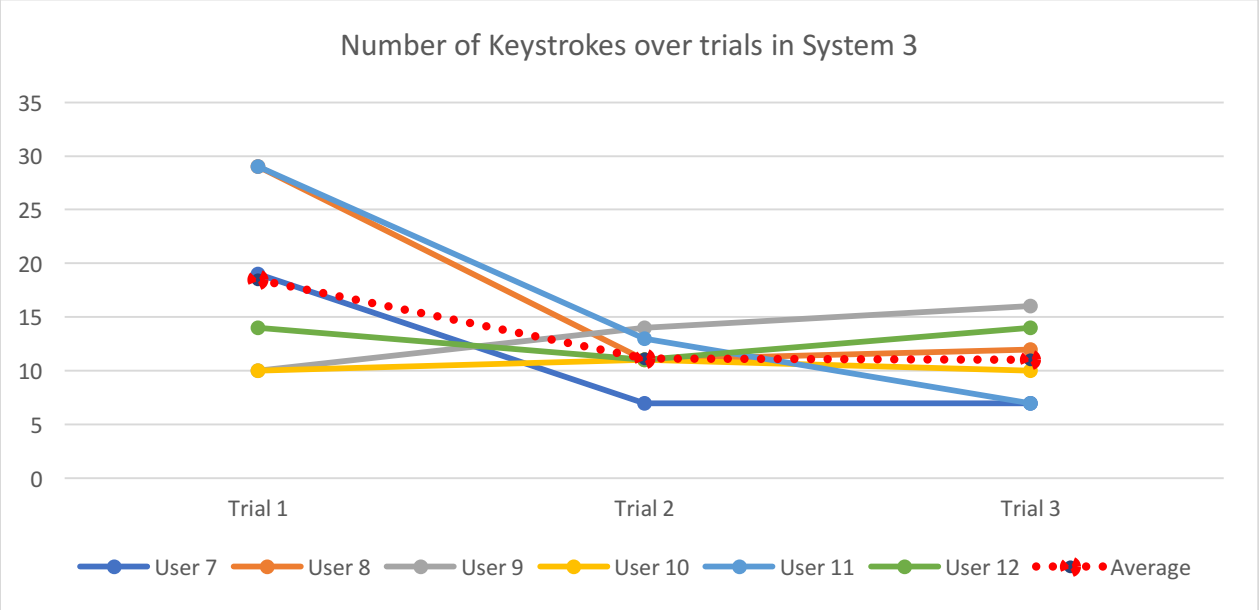

Figure 3.2: The number of Keystrokes in System 3 over the number of Trials

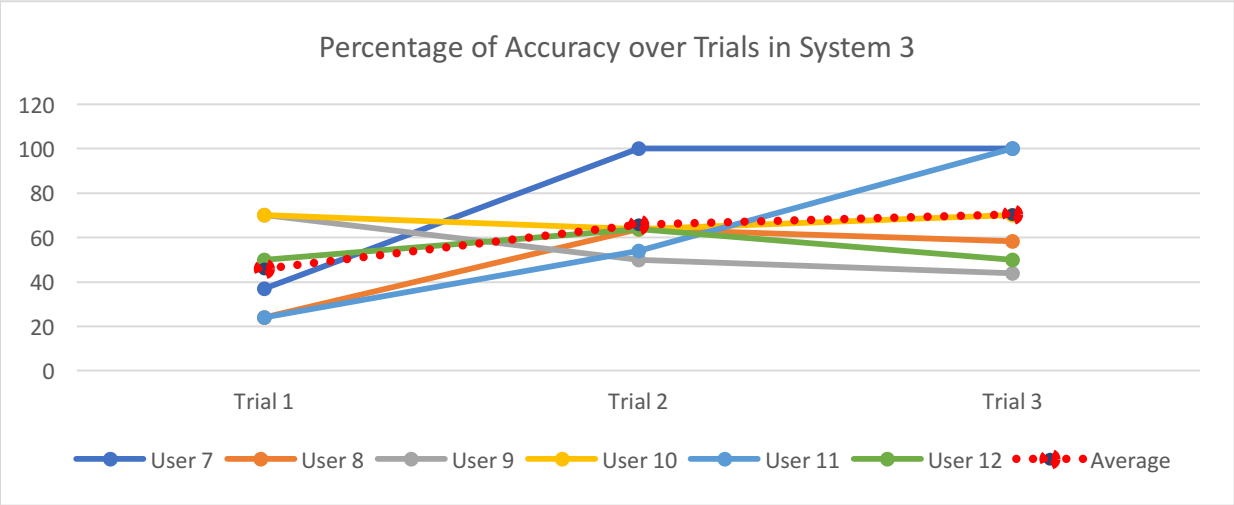

Figure 3.3: The percentage of Accuracy in System 3 over the number of Trials

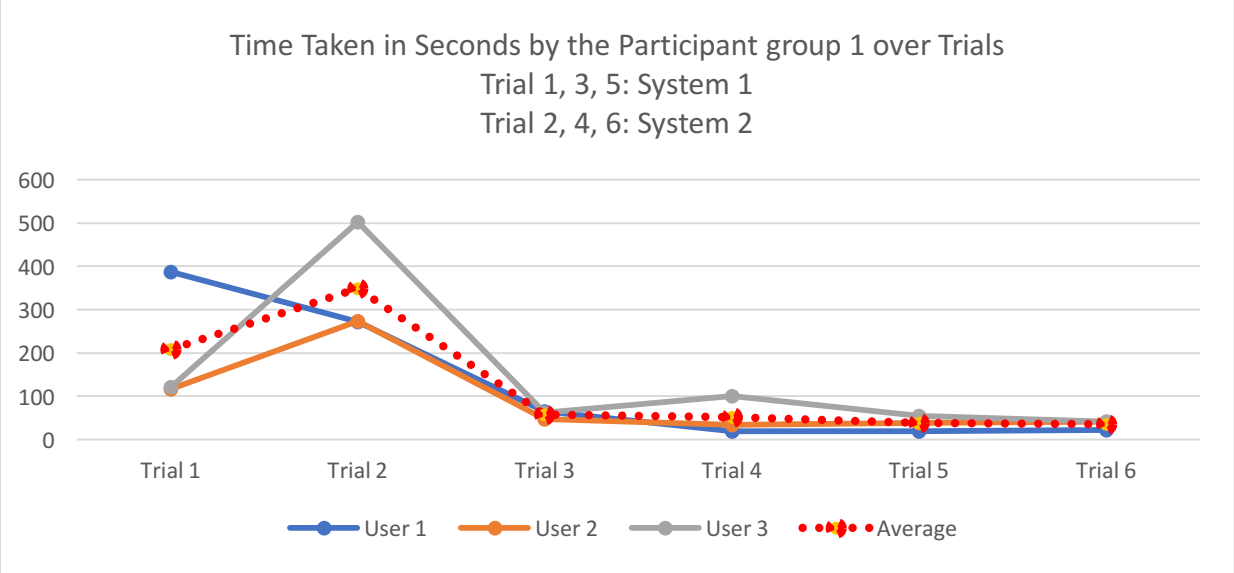

Figure 4.1: Time taken in Seconds by the participant group 1 with System 1 and System 2 over the number of Trials

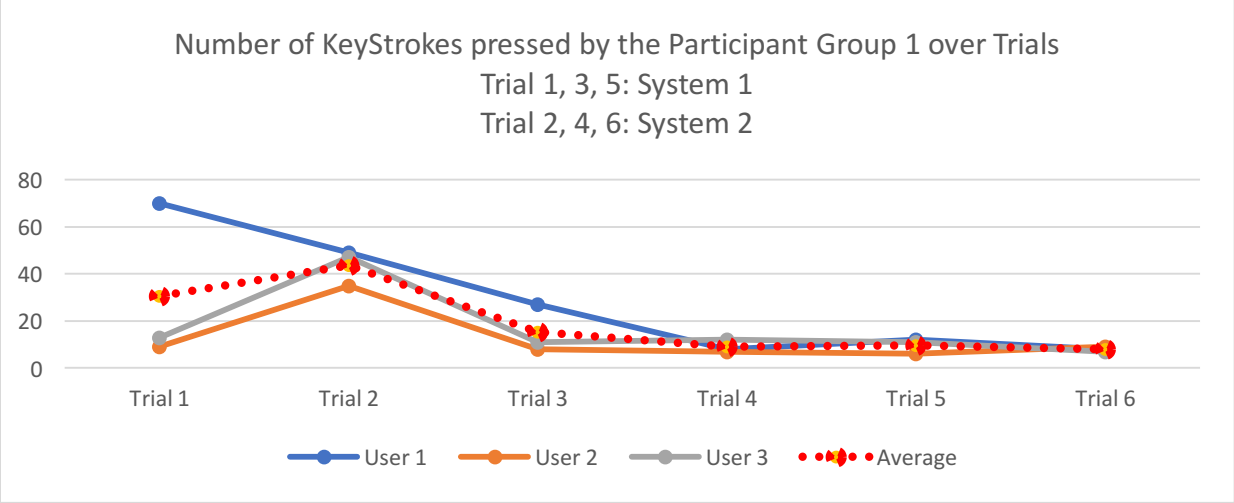

Figure 4.2: The number of keystrokes pressed by the participant group 1 with System 1 and System 2 over the number of Trials

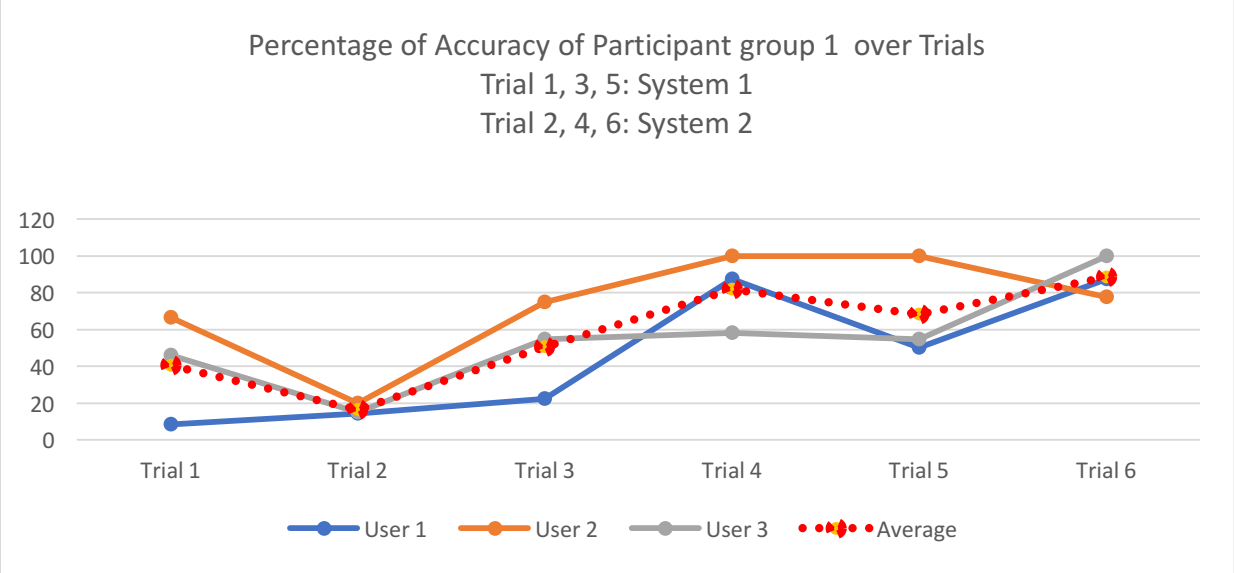

Figure 4.3: The Percentage of Accuracy of the participant group 1 with System 1 and System 2 over the number of Trials

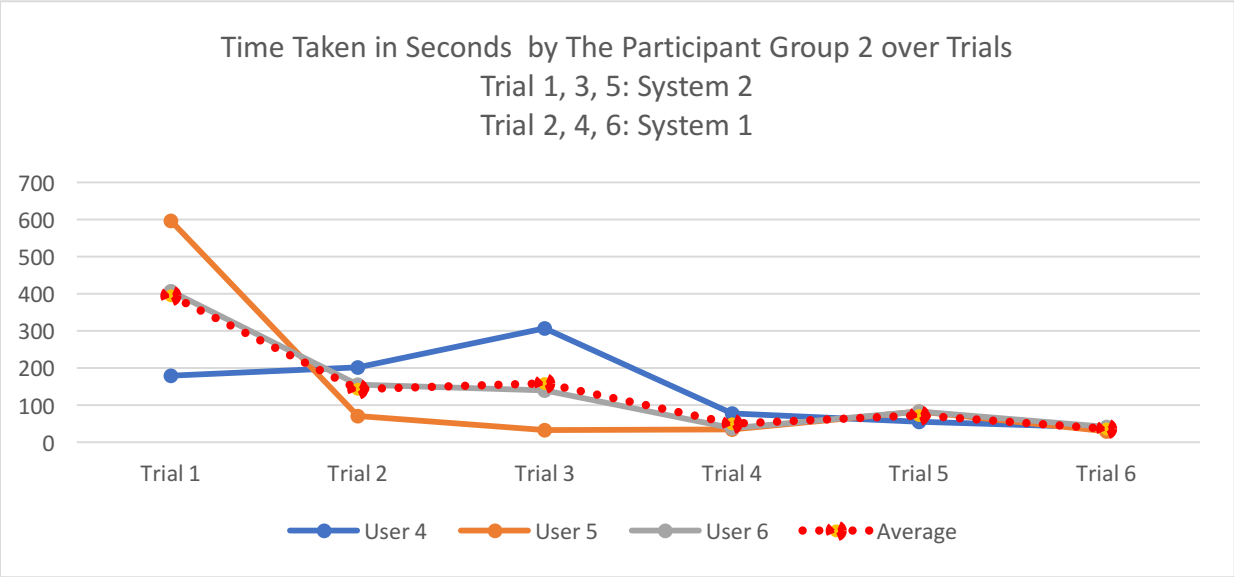

Figure 5.1: Time taken in Seconds by the participant group 1 with System 2 and System 1 over the number of Trials

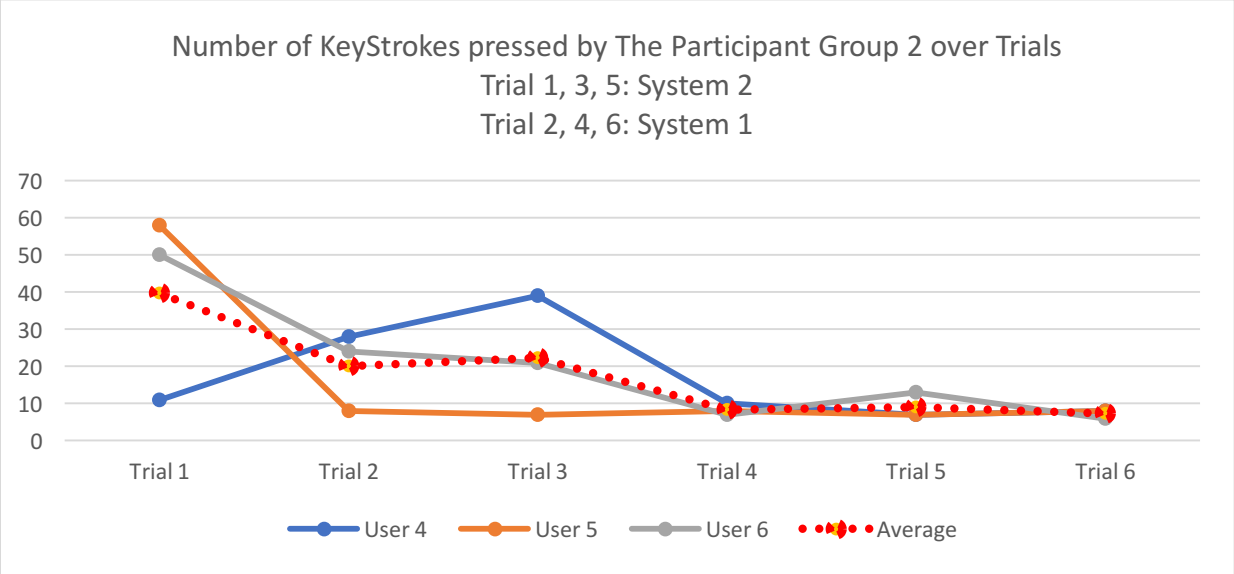

Figure 5.2: The number of keystrokes pressed by the participant group 2 with System 2 and System 1 over the number of Trials

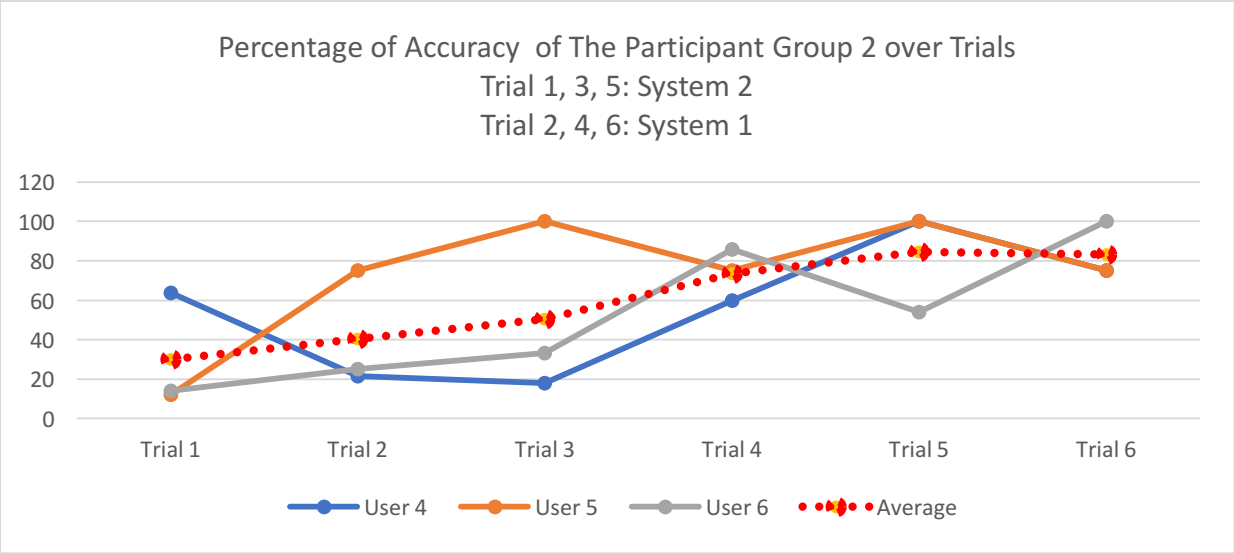

Figure 5.3: The Percentage of Accuracy of the participant group 2 with System 2 and System 1 over the number of Trials

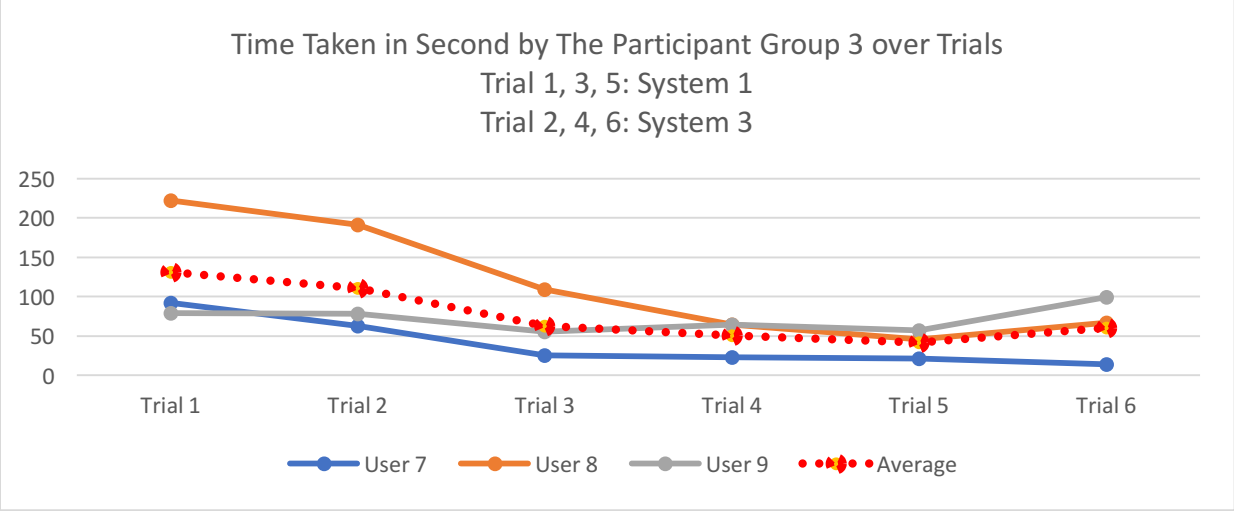

Figure 6.1: Time taken in Seconds by the participant group 3 with System 1 and System 3 over the number of Trials

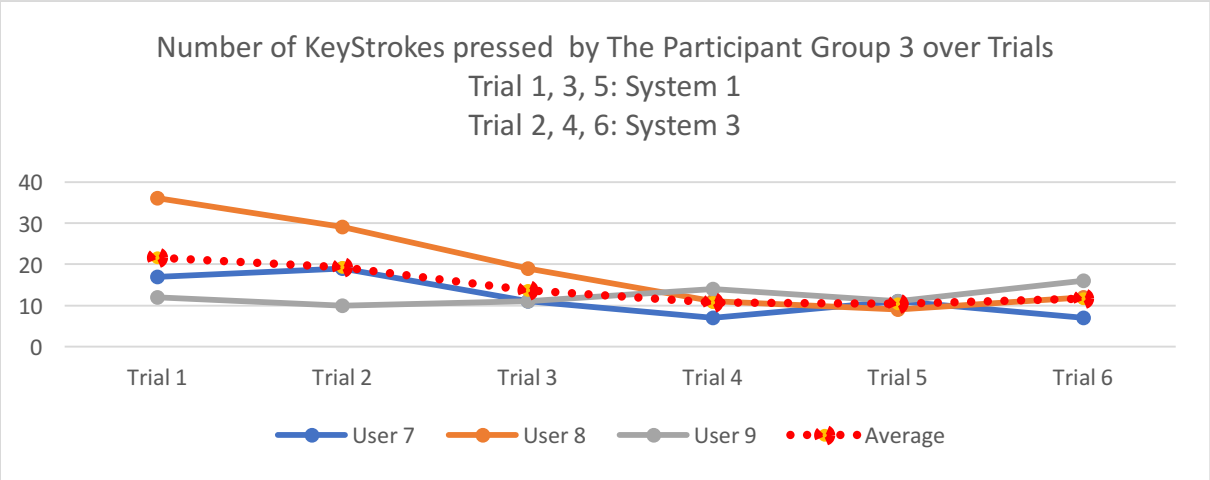

Figure 6.2: The number of keystrokes pressed by the participant group 3 with System 1 and System 3 over the number of Trials

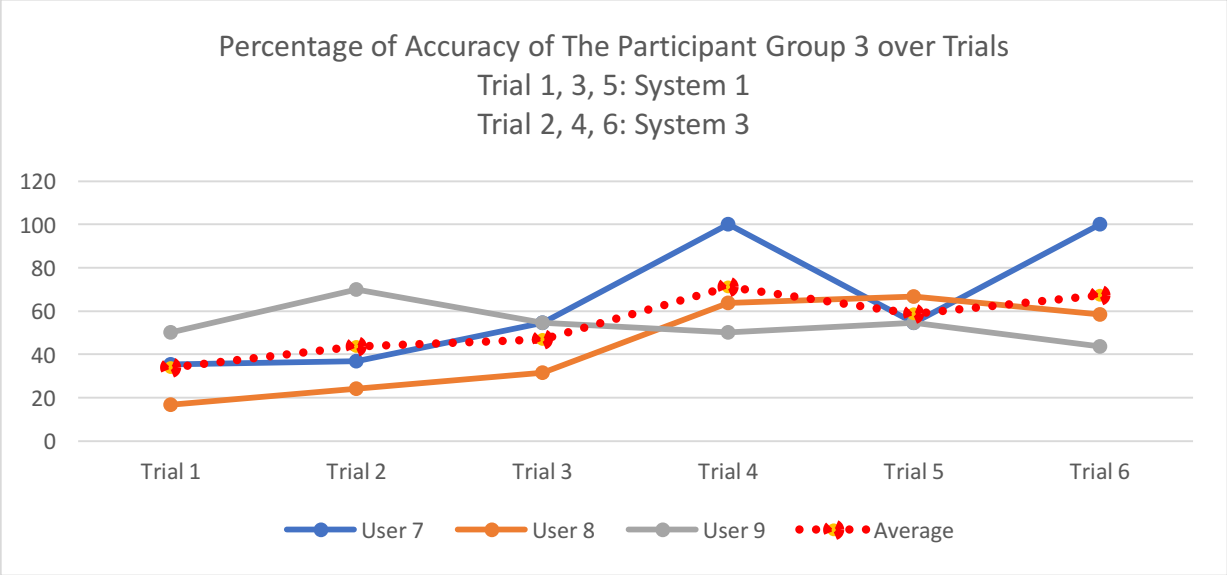

Figure 6.3: The Percentage of Accuracy of the participant group 3 with System 1 and System 3 over the number of Trials

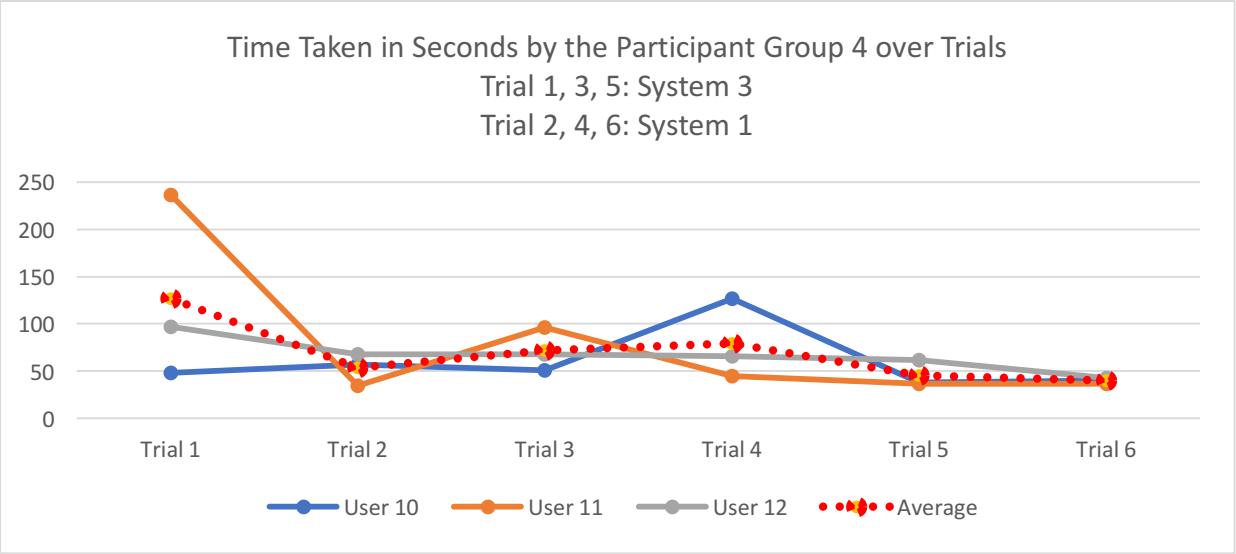

Figure 7.1: Time taken in Seconds by the participant group 4 with System 3 and System 1 over the number of Trials

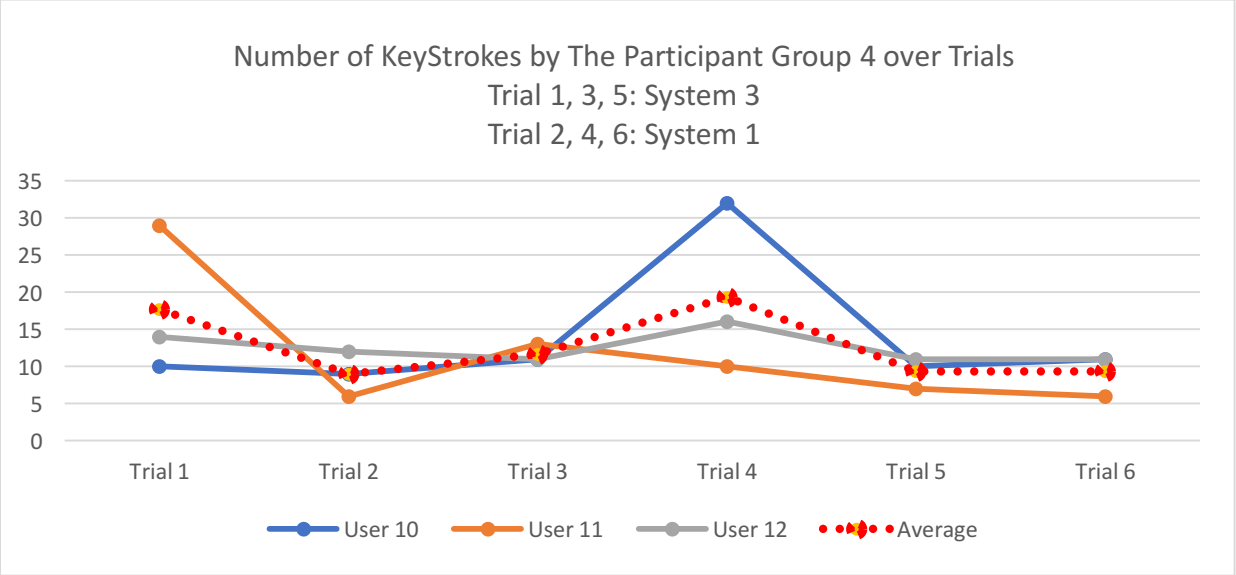

Figure 7.2: The number of keystrokes pressed by the participant group 4 with System 3 and System 1 over the number of Trials

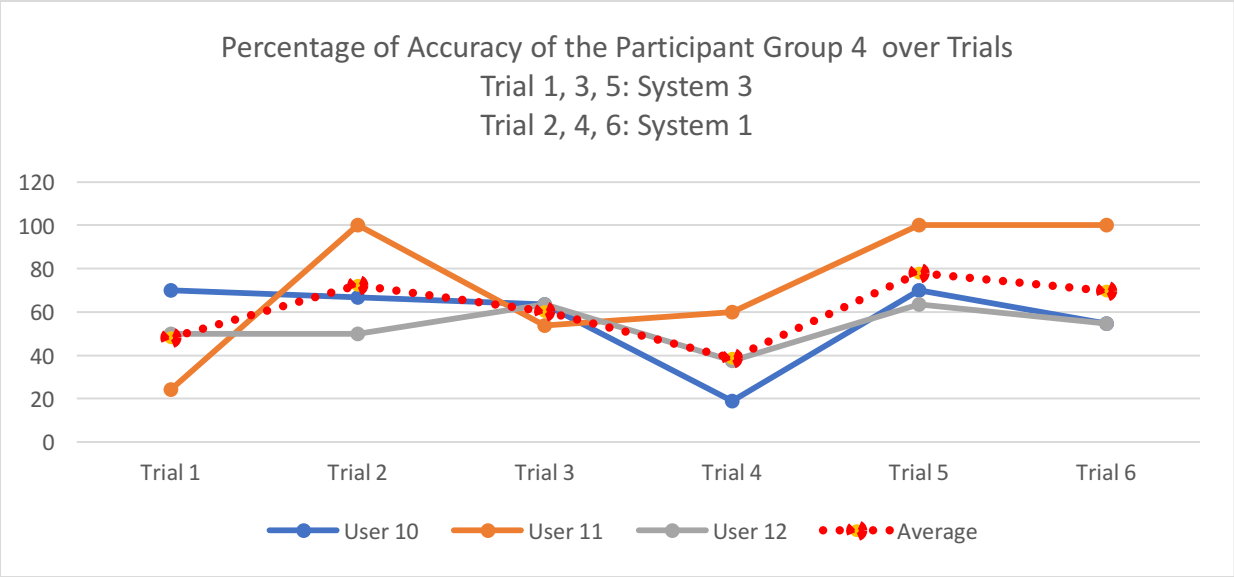

Figure 7.3: The Percentage of Accuracy of the participant group 4 with System 3 and System 1 over the number of Trials

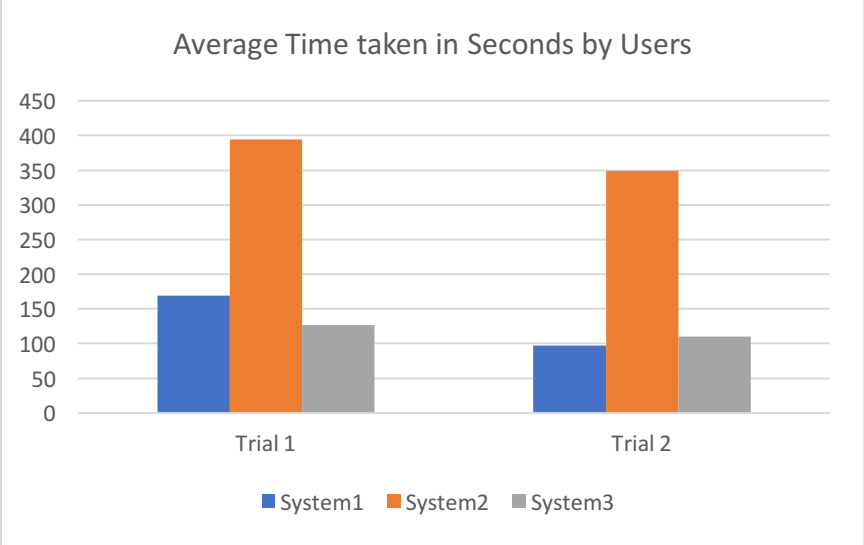

Figure 8.1: Average Time taken in Seconds by the participants with the Systems in Trial 1: when they first use the system and in Trial 2: when they have already experienced with another system once

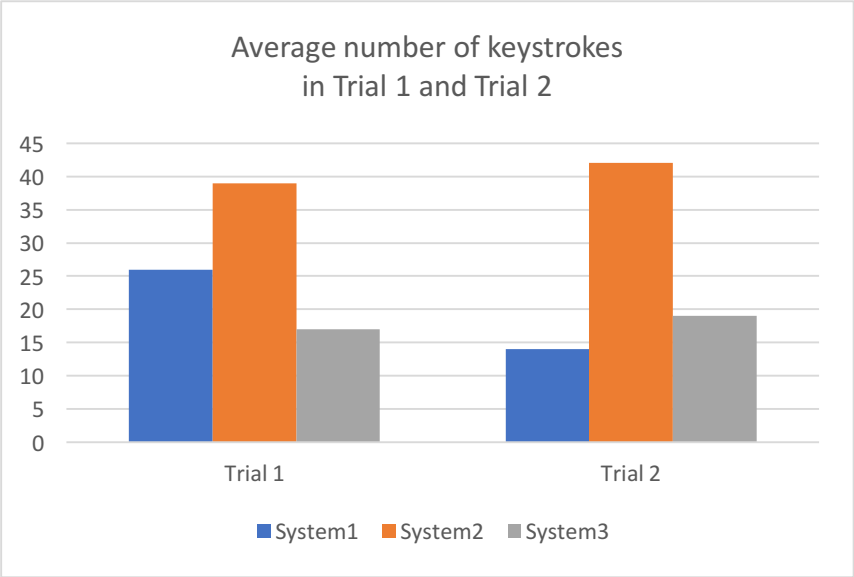

Figure 8.2: Average number of keystrokes pressed by the participants with the Systems in Trial 1: when they first use the system and in Trial 2: when they have already experienced with another system once

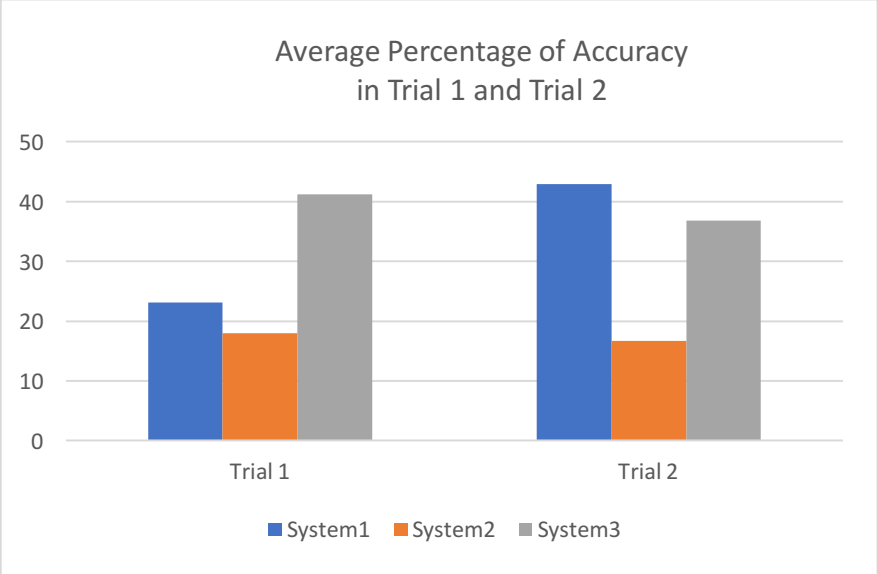

Figure 8.3: Average percentages of accuracies of the participants with the Systems in Trial 1: when they first use the system and in Trial 2: when they have already experienced with another system once

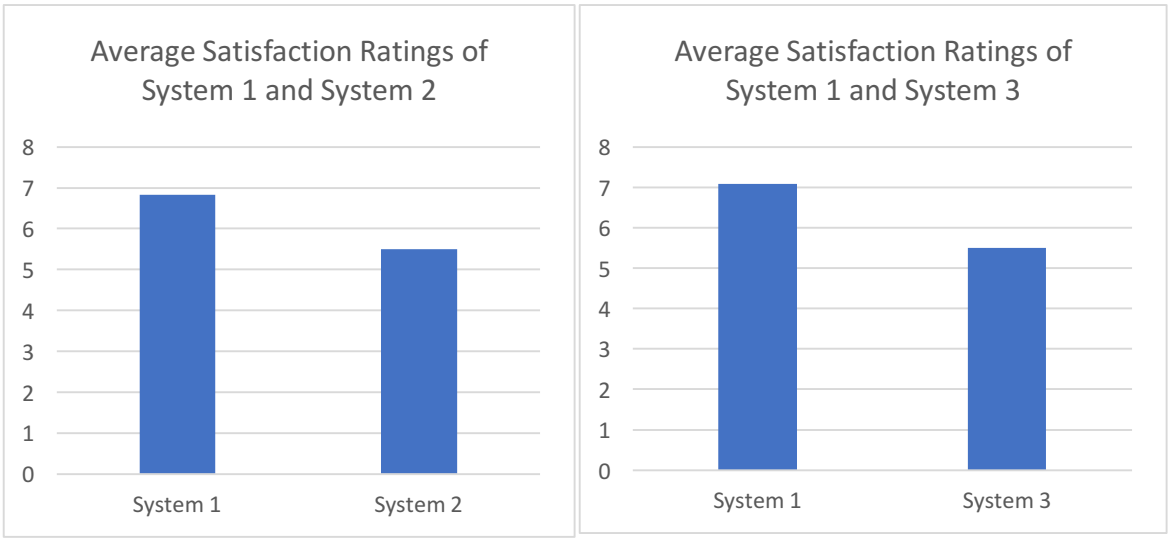

Figure 9: Average Satisfaction ratings of the participants with the System1 & 2 and System 1 & 3

4. Discussion

After each study was completed, each participant was asked a set of debriefing questions about the study and when asked if the system always responded as expected, I have got three negative answers from the participants about System 2. One said that <K> key in System 2 did not scroll forward as it said in help message. Another user said that he was stuck for a while and none of <J>, <K>, <Space> keys worked as they were supposed to do (this happens when the system is in help mode). "The functionalities of the keys in System 2 are confusing", two users mentioned. One user said that using System 1 helped him to learn about System 2. However, when the users used System 3 first and then System 1, two users have found System 1 is confusing as it lets user scroll through the read items of the chapter (when user presses <L> for help in the "Continue To Read" mode, the help message tells him or her how he/she can scroll through the read items).

During the study on System 1 and 2, when asked if one system was easier/harder to use and if they have preferred one system over another, I have got a variety of answers from them. one participant said system 2 was easier to use, one said both are almost same as he was stuck in help mode in system 2 and on the other hand, he had to scroll through the chapter names which caused him to forget to do the given task- select chapter two and play its content. Four participants said system 1 is easier. When asked which system they preferred between System 1 and 2, four participants said they preferred system 1 whereas two said they preferred system 2. In terms of system 1 and system 3, I found half of the participants liked system 1 and the other half preferred system 3 over

system 1. I also have asked to give a rating as per their satisfaction level in using the systems and Figure 9 depicts the rating they gave as per their satisfaction level when they were comparing System 1 with System 2 and System 3. Interestingly, the participants were more satisfied using System 1 than System 2 and System 3.

Figure 1.1 to 3.3 in the Discussion Section illustrates the learnability and practice effect of the same systems- how practice with one system helps the user to perform better in the subsequent trials with the same system. Figure 2.1 shows that system 2 has given the highest practice effect than the other systems in Figure 1.1 and 3.1 since almost all users have struggled in System 2 for the first time. And with repeated trials, users learned how to avoid the help mode and that caused this high learning effect. Moreover, the average number of keystrokes in the first trial with system 2 was more than double than with System 1 and system 3.

The performance of each participant group with different experiment settings are shown in 4.1 to 7.3. It appears from these data that almost all interface helped users to perform better with another system in the subsequent trial except the cases when the subsequent trial was with System 2. System 1 has better usability than system 2 and the users, on average, performed the tasks faster and more efficiently with system 3 than 1. Three of the participants got frustrated with system 2. Due to these results, it would appear that the abundant of auditory guidelines/instructions in every step of system 1, and so it is superior to system 2 in terms of the learnability. On the other hand, system 3 did not let the user scroll through the chapter names and it also has an auditory guideline on how to quit from the help mode and then how to select a book/chapter instantly and these features helped users to perform better than any other systems.

Figure 8.1 and 8.3 analyze the comparison of the inter-system practice effect more clearly. It shows that there was no learning effect of any system on system 2, users have almost always struggled with that system in both their first and second trial. However, System 3 has shown quite higher accuracy in both trial 1 and trial 2.

All these systems have some very good design decisions and some bad design decisions that I get to know after the experiment. System 1 has a logical and straightforward flow of functionalities- select book, select chapter, and continue to read. Any time the user can press L to listen to the help message that plays the functionalities of the keys. However, this system has an unnecessary mode “Content Not Available”. As the task was to play chapter 2, there was no necessity of having the option to select other chapters other than chapter 2. System 2 does not have this “Content Not Available” mode, user can select book and chapter 2 directly. However, it has a help mode and that help mode changes the functionalities of all keys and that caused frustration among users because there is no clear instruction on how to quit from the help mode. One user got so frustrated that he gave up the task of playing chapter 2 content and he quit from the program. In System 3,

there is also no “Content Not Available” mode but it also has a help mode. Nevertheless, this system has a clear auditory guideline on how to exit from the help mode and select a book. That helped the user to perform the task faster but easily. I opine that System 3 has met the usability goal because the design of this system helped users to use it efficiently. System 1 also has met the usability goals because of the sufficient number of auditory guidelines/instructions.

In general, system 1 and 3 did not hinder the user’s ability to complete the tasks. Users spent substantially less time learning the keystrokes and making mistakes with system 1 and system 3 than they did with system 2 and therefore they were able to complete the tasks quicker and more accurately. However, System 3 required fewer keystrokes, less time than system 1.

After analyzing the answers and collected data, I found a possible modification would increase the user satisfaction of System 1 and System 3. I believe, if System 1 would not have the mode “Content Not Available”, users would not spend much time to play chapter 2 and thus it would be the most usable system among all the three. And if we could remove the help mode from system 3 and add more auditory guidelines, that would cause fewer keystrokes and less time spent to perform the task.

5. Conclusion

In conclusion, system 1 and system 3 were more usable than system 2 in the sense that users were able to accomplish tasks faster and more accurately. This suggests that designing a simple system but with necessary functionality is a crucial element in building an interface, and to the average user’s experience. Having a poor guideline with more advanced features, on the other hand, will leave users unsatisfied, and often frustrated, as I have experienced during the study.