Abstract

The study investigates the usability two variations of soft keyboards – alphabetical keyboard and an Alpha-keypad. The results showed that alphabetical keyboard is a better method of input regardless of the level of expertise of the user. The high rate of errors in the keypad interface also suggests the effect of negative transfer from a conventional cell-phone text input mechanism.

1. Introduction

Text entry is conventionally done via input through a dedicated source – the keyboard. However, implementation of such hardware may not always be plausible for input into specialized systems, such as a fax machine, cell phone, or a kiosk. New technology has allowed text entry through a touch screen interface. However, the processing overhead, as well as the cost of such an input mechanism may not be feasible for mass consumer products. Applying a concoction of the two aforementioned input methods – a 'soft keyboard' displayed on screen with a mouse to press the keys on this virtual keyboard – may yield an effective mechanism to enter text into a system.

The main premise of this study is to test and analyze the usability of two different variants of a 'soft keyboard'.

1. Alphabetic Keyboard: A complete view of all the keys, accessible with a single mouse click.
2. Phone Keypad: All the alphabets are distributed over the, much smaller, 0-9 digit keypad, and requires multiple presses to obtain the desired alphabet.

Precedence is given to Alphabetic keyboards over QWERTY keyboards as it is assumed that most of the users are beginners with little/no prior typing experience . On the same note, a research conducted by Mackanizie, et al. (1999), QWERTY keyboards are more efficient than ABC keyboards for expert users; primary due to positive transfer from conventional use of a Personal Computer Keyboard. However, novice users were able to enter text at > 9.6 wpm – an indication that the chronologically sequenced keys reduce visual scan time, as all participants learnt the alphabets in the same order.

Phone keypads are considered to be another viable option as cell phone use is common amongst our target users and the possible positive transfer effect of this interface makes it a good candidate for our study.

Other alternatives such as handwriting recognition, voice-to-text are dismissed due to poor recognition rates (less than 70%).

2. Description

The initial research was carried to test and compare the usability of the two interfaces and empirically analyze the data to see if one is better than the other.

2.1 Participant Pool

The main demographic involved in the research were university students with Computer science backgrounds. This directly implies that the users were comfortable with using the equipment required to carry out the study, and had prior exposure interfaces similar to the ones being observed (though there are subtle variations).

2.2 Conditions

The users were asked to interact with the two different interfaces – the alphabetic keyboard, and the numerical keypad. Both interfaces required the user to click on a start button to begin and a finish button when they were done. The time taken and errors made were collected during this time window. Furthermore, the phone keypad required addition explanation on how to proceed to the next alphabet (by pressing “<” once) and executing a space between words (by pressing “<” twice).

2.3 Tasks

The users were assigned an interface at random and given three sentences to enter using the given interface. The first one was a test, to allow the user to get familiar with the system they are working with. The second and third sentences were ordered according to increasing difficulty. There were pre/post experiment questionnaires to obtain qualitative data and any feedback the user had.

2.4 Design

The experimental design can be described as a 2 x 2 with two levels of expertise (between subjects; novice and expert) and two types of interfaces (within subjects).

2.5 Procedure

Users were given a set of questions before the experiment to assess their familiarity with the interfaces and their level of expertise. The participant was then given a sheet with three sentences and one of the two interfaces to enter the given sentences. The users were given an overview about how to proceed and how to finish once they were satisfied with their input–first by the experimenter by typing, “test” and then letting the user type in the first test sentence. After the users had completed the entry, they were given the second interface and asked to repeat the process. After completion, the users were asked a few questions to obtain any qualitative data on preference and opinions.

2.6 Apparatus

The experiments were carried out at the HCI laboratory. The main component was a Windows based computer with a compatible mouse. Furthermore, the users were seated and were asked to perform the tasks in a natural environment – complete with white noise and/or other distractions.

2.7 Independent and Dependant Variables

The two dependent variables that we intend to measure are:

- The time taken for the participant to enter the given text
- The number of errors made while entering the given text

There are a number of independent variables that may influence our results. These include the level of expertise, familiarity to the system being tested, mouse sensitivity, effects of the surroundings etc.

2.8 Hypothesis

H_a: One user interface is better than the other regardless of the level of expertise of the user

There are limitations to the experimental design. It may introduce nuisance/confounding variables. The demographic we are testing does not reflect the general user group. All the participants are Computer Science students and exhibit technical knowledge and experience. Furthermore, the setting was no intended for research (rather teaching) and it may introduce some unwanted influence into our data and results.

3. Results

Anova: Two-Factor With Replication

SUMMARY	Phone	Alphabetic	Total
H			
Count	9	9	18
Sum	1208	666	1874
Average	134.2222	74	104.1111111
Variance	1644.444	593.5	2013.163399
L			
Count	9	9	18
Sum	1148	614	1762
Average	127.5556	68.22222222	97.88888889
Variance	1375.528	547.4444444	1836.810458
Total			
Count	18	18	
Sum	2356	1280	
Average	130.8889	71.11111111	
Variance	1432.928	545.751634	

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	348.4444	1	348.4444444	0.334969	0.5668	4.149097
Columns	32160.44	1	32160.44444	30.91669	3.89E-06	4.149097
Interaction	1.777778	1	1.777777778	0.001709	0.967281	4.149097
Within	33287.33	32	1040.229167			
Total	65798	35				

Table 1.1 –Two-way ANOVA test

t-Test: Paired Two Sample for Means

	Alphabetic	*Phone*
Mean	70.61904762	128.047619
Variance	477.547619	1275.547619
Observations	21	21
Pearson Correlation	0.709852336	
Hypothesized Mean Difference	0	
df	20	
t Stat	-10.36188018	
P(T<=t) one-tail	8.6722E-10	
t Critical one-tail	1.724718218	
P(T<=t) two-tail	1.73444E-09	
t Critical two-tail	2.085963441	

t-Test: Paired Two Sample for Means

	Alphabetic	*Phone*
Mean	0.4	2
Variance	0.3	5.5
Observations	5	5
Pearson Correlation	0.194625	
Hypothesized Mean Difference	0	
df	4	
t Stat	-1.55406	
P(T<=t) one-tail	0.097569	
t Critical one-tail	2.131847	
P(T<=t) two-tail	0.195138	
t Critical two-tail	2.776445	

Table 1.2 – speed t-test for all observations Table 1.3 – Error t-test for subset of observations

4. Discussion

The analysis of the obtained data shows that the two tailed t-value for the error rate is calculated to be 2.78 (Table 1.3). Using the degrees of freedom and inferring the probability from the t-test table, we can calculate the probability of the null hypothesis being true to 0.05 or 5%. Conversely, there is a 95% chance that our alternate hypothesis is true. It is can be fairly assumed that there is a difference in the usability of the two systems based on the number of errors made.

Similarly, applying the same ideas on our data for the time taken (Table 1.2), we had initially obtained a t critical two-tail value to be 2.086. Given the degrees of freedom to be 20, we see that the probability of the null hypothesis being true as 0.05 or 5%. This implies that there is a difference in the rate of alphabet entry between the two interfaces.

Since we can now assume, with a fair accuracy that there is a difference, we need to assess which of the two interfaces is better. Furthermore, we wish to remove the level of expertise from our analysis to obtain empirical evidence about which system is better. The results of the ANOVA are given in table 1.1. Looking at the F-value between the columns (i.e. the two interfaces in question), the ratio of 30.92 suggests a significant difference. The f-ratio is large because we had a small sample size. It would decrease if we increase the sample space. Furthermore, we see an interaction between the two interfaces – Alphabetic interface is better than the keypad interface, regardless of the expertise level in subjects.

It may seem counter intuitive that there is not a significant effect of positive transfer on the keypad input; and it is less effective regardless of the user's experience. This may be attributed to the subtle differences in the interaction. For example, the user had to click '<' twice to proceed to the next alphabet. This is not the case in regular cell phones as the system automatically recognizes a jump if a different key is pressed, or there is a delay between the alphabet selection. We may infer that there was some negative transfer due to the variation in the conventional system and our keypad. The effects of the 'Power of Practice' are assumed to be neutralized because of randomization of trials. However, the sample space may not be reflective of the actual demographic as most users are technology savvy and familiar with effective mouse use.

The results agree with the research conducted by Mackanzie et al. (1999); the keypad entry yields the lowest wpm (words per minute). The research also showed that QWERTY could significantly better than the Alphabetic keyboard due to the distance between alphabets that make up a word. However, given our scope of the experiment, Alphabetic keyboard is a more appropriate option than the Keypad option.

More research may be able to shed light on the effect of negative transfer on the keypad interface, and if it yields a better system if such effects are mitigated. It may also be more pertinent to conduct a comparative study on QWERTY vs. Alphabetic keyboard to assess their usability in the light of Fitt's Law.

4. Conclusion

The results conclude that the Keypad interface is slow and more prone to error. Comparatively, alphabetic keyboard is better and yields faster inputs with fewer errors. However future studies may be able to indicate if Alphabetic keyboard is universally better for new users, if there are negative transfer effects from QWERTY users – in the context of visual inspection time to look up alphabets, and the input accuracy associated to each interface in terms of Fitt's Law.

Reference:

I. Scott Mackanzie, Shawn X. Zhang, and R. William Soukoreff. *Text Entry using soft Keyboards*, Behavior & Information Technology, 1999, Vol.18, No. 4, pp. 235-244