Original Research Article

Evaluating Soft Keyboard Layouts by Using Paper Mockups

ABSTRACT
Two experiments to compare text-entry were carried out using Soft Keyboard Layouts. Paper mockups were used and we carried out manual timing in a relaxed situation, even though within a classroom environment. Volunteer students were paired and made to work in such groups, serving as participants and also helping out in the timing process. One set of students served as participants at first and the other timed them. They then switched roles halfway and exchanged the tasks of participation and timing. Participants tapped the phrase *"the quick brown fox jumps over the lazy dog"* on each keyboard layouts five times using stylus. The time for word entry was measured and used to determine the speeds; which were found to be 24.8 to 34.3 for QWERTY Keyboard Layout and 11.9 to 14.8 for Opti Layout. We discussed the advantages and shortfalls of the methods of evaluation.

Keywords
User Interface, Soft Keyboard Layouts, text entry, Paper mockups, Participants, Interfaces

1. INTRODUCTION
One of the most popular areas of research in the mobile communications world today concerns developing efficient text entry methods. Most online businesses are done on mobile devices today in a phenomenon known as Mobile Commerce (M-Commerce) [1,2]. Customers and other users of the ever increasing mobile devices become impatient if anything hampers their quick access to information. And because they have to also send in requests as quickly as possible, they need interfaces on these devices that allow for quick data entry [3,4]. Thus, the subject of developing efficient entry method is now a sort after area of research. The primary concern of this paper is to formerly develop, test and critique a rapid evaluation method for new techniques of input in the mobile ecosystem.

1.1 Mobile Text Entry
The proliferation of mobile devices comes with its unique challenges. While users are happy with the mobile nature of the devices, there is a price to be paid by them for the mobility. The user interface providing the physical means of input and other interactions with the device are constrained by size, being so small that full sized input devices like keyboards and mice are not practicable in them. For this reason, other forms of input designs are required. These could be best with such inputs like speech, facial sensing and recognition, interfaces with fewer and smaller keys, and so on. The use of stylus-based mobile systems supports gesture recognition and also tapping. One common use of stylus-based tapping is on soft keyboards found today on most mobile devices. The soft keyboards are easily implementable and afford users a good alternative to handwriting. Apart from the QWERTY and Phone layouts, other physical layouts are becoming quite interesting to users today. Such Alternatives as Dvorak Lewis (1997), cited in MacKenzie1 and Read (2007) or chord keyboards could support higher rates of data entry; though, proficiency can only be achieved with a lot of practice. This, together with a huge coverage area for QWERTY, ensure that QWERTY continues to be seen as the keyboard commonly embraced for desktop computing. There isn't as much arguments for QWERTY in terms of soft keyboards. The device is virtual, instead of physical, and so renders manufacturing costs to the software, which are but one-time. Hence, investigating soft keyboard layouts design space is now an important area of research.

1.2 Users Expertise
According to Hughes et al. (2012), as cited in MacKenzie1 and Read (2007), researchers or developers design of text input methods focusing on the potential or expert entry rate of a design. Although the novice experience should be treated as principal, new text input methods has got to succeed. This is due (at least, partially) to the target market. Before now, mobile devices such as cell phones and other Personal Digital Assistants (PDAs) were known to be specialized tools for professionals. Today, the case isn't the same as they are freely and easily accessible and used by consumers. This means that usability is important either as immediate (for experts) or walk-up (for novices coming up). In other words, an entry design could be seen almost as of little or no use if a user needs to put in prolonged practice in order to establish expert text entry rate. This is because, the initial experience and frustration of consumers may discourage them and they may never put in the time and effort required to become experts after all.

1.3 Evaluation
It is rather time consuming to carry out Empirical Evaluations of new interaction techniques as it is also labor-intensive. So, another research topic closely related to this is the development of efficient methods of evaluation. A lot of different methods exist which are in use today. It is clearly efficient, as implementation is deferred until evidence on problems in the interface is gathered. There is the provision of another convenient and efficient means to gather feedback from users in paper mockups. In this case, we implement an interface on paper and user impressions are besought, possibly across numerous hypothetical implementations. In as much as such methods are standard and successful, prior work with paper mockups is solely qualitative. We are interested in exploring the use of paper mockups for quantitative evaluations which we see as important since entry speed is the major concern of research questions pertaining to text entry on computing

devices. We increased the efficiency of the evaluation by simultaneously testing our participants, and engaging them also as experimental assistants in the exercise. We conducted two experiments from which we tested our methods, results, and analyses.

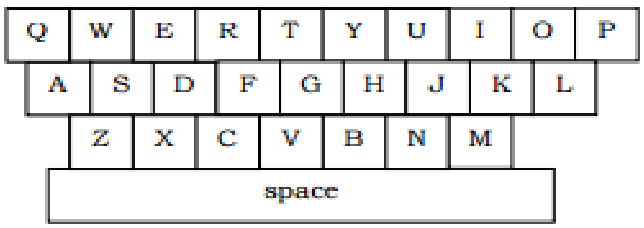
2 Method – Experiment 1

2.1 Participants

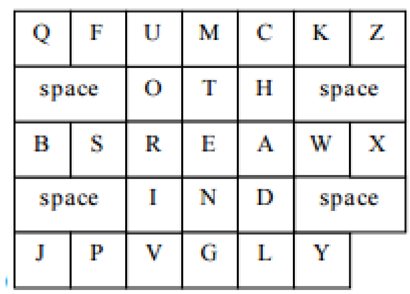
We drafted twelve volunteer participants of university age from among our friends in the local University. These were students enrolled in different courses but who utilize the University computers and network for their entire academic work on campus. The recruited students served both as participants and as assistants, helping out in timing while members of the other group tap on the keyboard mockups.

2.2 Apparatus

We selected the QWERTY and Opti soft keyboard layouts for this evaluation. Opti is a great layout MacKenzie (1999), as cited in MacKenzie1 and Read **(**2007**)** which was designed using the Fitts-digraph model of Soukoreff and MacKenzie. We used predicted entry rates of 30.0 words per minute (wpm) for QWERTY and 42.2 wpm for Opti MacKenzie (2002), as cited in MacKenzie1 and Read **(2007)** for experts. The layouts were both implemented as paper mockups as shown in Figure 1 below.



A. QWERTY Keyboard Layout(image)



B. Opti aaKeyboard Layout(image)

Figure 1: QWERTY & Opti Keyboard Layouts

The paper mockup keyboard layouts measured 10.0cm by 4.0cm for QWERTY and 7cm by 5 cm for Opti. Even though these dimensions are greater than normal soft keyboards sizes on most PDAs we expect that this would not adversely affect the performance considering that there are both theoretical and empirical evidences MacKenzie et al. (2001) as cited in MacKenzie1 and Read (2007) that text entry rates for soft keyboards are not affected by the size of the layout, within reasonable limits. Our greatly simplified apparatus didn't allow for electronic measurement of entry times due to the absence of sensing features or other electronic experimental software. We therefore measured entry times using mobile phones in timing mode and recorded the results by hand.

2.3 Procedure

We requested participants to read and memorize the 43-character phrase bellow, then tap them on the paper mockup keyboards:

"the quick brown fox jumps over the lazy dog"

Styluses were used to tap the characters of the phrase on the soft keyboard layout. Each participant used a tip-covered pen as stylus so as to allow the layout sheets remain tidy all through the exercise. Participants were instructed to quickly enter the phrase and at the same try to avoid mistakes. Because we generated no text and recorded no accuracy, we made further elucidation to participants on the need to tap in the words quickly but not recklessly; and for them to accurately tap the right keys on the soft keyboard. Participants were grouped in two: a group tapped while the other monitored and measured the entry time. When the timer said "start" the participant started tapping. The participant continues to tap and when he was about to tap the last letter (which if "g" from the last word dog) he also shouts "STOP". This is to enable the timer know where to stop since text were not generated electronically on the soft keyboard layouts on paper. The entry time was taken and written in seconds in a record book. The process was repeated five times for each layout. In the first five times, the participants in the first group did the tapping while those of the second group timed them. Then they reversed the procedure and had the second to do the tapping while the first group timed, also for five times. Each group comprised of six participants both for entry and for timing. The experiment was conducted in a classroom during lecture break within a time period of about one hour.

2.4 **Design**

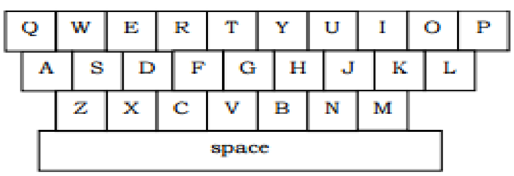
The experiment had a mixed design of 2 x 2 x 5 treats. Two Groups A and B; each with six participants and constituted one level of between-subjects factor while the Keyboard Layouts formed two levels of within-subject factors comparing QWERTY and Opti, giving with five levels of trial (levels 1 until 5). With six participants in 2 groups we had 12 participants each running through the procedure five times on each keyboard layout and giving us a total of 120 phrases tapped on the two keyboard layouts. 6 (participants) x 2 (groups) x 2 (layouts) x 5 (trials) = 6x2x2x5 = 120 phrases. The only behavior that was measured was the phrase entry time for each participant per group. The interest was to get the rate at which words were entered in the different keyboard layouts. This is obtained by number of words per minutes. So, the number of characters in the phrase was divided by the number of words to determine the average length of words. This was found to be approximately 5. Entry speed was therefore computed as $r = (43 / 5) / (t / 60)$, where r = the required speed, 43 = total number of characters to tap in the phrase, 5 = number of characters per word, t = the recorded entry time in seconds, and 60 stands for the number of seconds that make up a minute.

2.5 Results/Discussion – Experiment 1

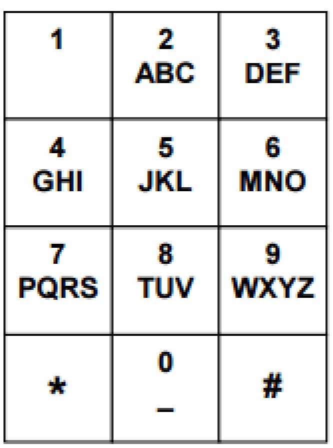
Group effect and interactions were statistically insignificant which showed that we achieved the desired result by the order of layout testing. The mean entry speed for the QWERTY layout was 26.4 wpm while that of Opti was 12.0; giving an overall mean of 19.2 wpm. The difference was statistically significant ($F_{1,10} = 797.0$, $p < .0001$). There was substantial difference by participant. For QWERTY, participant means for five trials were between 18.6 wpm to 30.5 wpm while those of Opti stood between 6.4 wpm and 16.2 wpm. It could be seen thus, that participants took the exercise with varying level of seriousness; considering speed and accuracy. 33.5 wpm was the highest speed recorded in single phrases for Qwerty 20.7 wpm was recorded for Opti. A significant effect for Trial also existed ($F_{4,40} = 50.7$, $p < .0001$), which means that participants increased entry speed with practice. In the same vein, there was significant Layout by Trial interaction ($F_{4,40} = 2.7$, $p < .05$), even though this was less than the main effect.

3 Method – Experiment 2

The second Experiment was carried out in the same way as the previous, the difference occurring in the fact here the QWERTY keyboard Layout was compared with a Phone Keyboard Layout whereas in the first, the comparison was between the QWERTY and Opti keyboard Layouts. The same category of participants was utilized in this as in the first experiment without repeating any of them. The different keyboard layouts are shown in figure 2 below:



C. QWERTY (image)



D. Phone Keyboard Layout (image)

Figure 2: QWERTY & Phone Keyboard Layouts

The phone keypad could be confusing for text entry so, participants were further instructed to tap each key one time only for any character they wished to tap whether or not the character they want to tap appears at the beginning on the key.

3.1 Results/Discussion – Experiment 2

Much like the first Experiment, the second Experiment showed insignificant effect and interactions for Groups. QWERTY keyboard layout entry speed was 45.8% higher than the Phone keypad layout. Its mean was 27.7 wpm while that of Phone keyboard was 19.1. The difference was statistically significant ($F_{1,22} = 65.80$, $p < .0001$). The main effect for Trial was also significant ($F_{4,88} = 54.22$, $p < .0001$) which means that participants improved their entry speeds with practice.

4 Summary

A general critique of the methods of experiment is provided in this section; examining procedure in the two experiments and comparing the results therefrom.

4.1 Combined Results

We compared two important points: the two QWERTY keyboard tests in both experiments with the Opti on the one hand and with the Phone keypad layout on the other. For QWERTY, the means were 26.4 and 28.5 wpm. In generally, it wasn't surprising to have the "between study differences" here as the participant differences have their individually different experiences with the usage of Computers and also of these mobile devices.

4.2 Validity of Comparisons

The research intended to test a simple method for evaluating soft keyboard layouts. The results showed that the goal that was met in this comparison is that of simplicity, as there wasn't enough scrutiny to this experiment when it comes to comparison with others wherein apparatus and procedures are more realistic and thorough. The QWERTY layout gave results that are close in the two experiments while the Opti and Phone layouts gave results that were lower.

4.3 Critiquing the Method

The results of our experiment are reasonable to some extent, being limited in the sense that only one variable was dependent in the whole exercise. Also, the fact that accuracy wasn't measured in the recording the text entry time undermines the accuracy of the entire experiment. The experiment design and implementation had the advantage though, of being easy as participants only needed to use simple tools in the mockup keyboard and not to any actual physical device. And this was achievable too within one hour and doesn't have to take too much time of the participants.

4.4 Accuracy in Measurement

In this experiment accuracy could not be measured due largely to the fact of the virtual apparatus used and the fact that no characters were printed while they were tapped onto the paper mockup keyboards. It is therefore possible in future experiment to adapt better procedures that would take into consideration the accuracy and eliminate errors as much as possible. This could warrant participants to use pens that would show the marks of their tapping on the paper. In this way, it may be necessary to provide one paper mockup keyboard for each tapping of the phrase.

5 Conclusions

We have established that paper mockups and hand timing could be used to test the rate of text entry on soft keyboard layouts and the efficiency thereof. Participants were tested together and also assisted in the timing process of the experiment. We measured 26.4 and 28.5 wpm for QWERTY layout and 12.3 wpm for an Opti layout. If the evaluation could be made with more proper comparisons, their outcome would be seen to be more consistent, which means that we had a useful methodology which was also quick and efficient means to test soft keyboard layouts empirically.

Consent

As per international standard or university standard, Participants' written consent has been collected and preserved by the author(s).

COMPETING INTERESTS DISCLAIMER:

Authors have declared that no competing interests exist. The products used for this research are commonly and predominantly use products in our area of research and country. There is absolutely no conflict of interest between the authors and producers of the products because we do not intend to use these products as an avenue for any litigation but for the advancement of knowledge. Also, the research was not funded by the producing company rather it was funded by personal efforts of the authors.

7 References

1. MacKenzie, S. & Read, J. C., 2007. Using Paper Mockups for - Evaluating Soft Keyboard Layouts. pp. 1-11.

2. McGill M, Brewster S, De Sa Medeiros DP, Bovet S, Gutierrez M, Kehoe A. Creating and Augmenting Keyboards for Extended Reality with the K eyboard A ugmentation T oolkit. ACM Transactions on Computer-Human Interaction. 2022 Jan 17;29(2):1-39.

3. Hu Y, Wang B, Wu C, Liu KR. mmKey: Universal Virtual Keyboard Using A Single Millimeter-Wave Radio. IEEE Internet of Things Journal. 2021 May 27;9(1):510-24.

4. Fennedy K, Srivastava A, Malacria S, Perrault ST. Towards a Unified and Efficient Command Selection Mechanism for Touch-Based Devices Using Soft Keyboard Hotkeys. ACM Transactions on Computer-Human Interaction (TOCHI). 2022 Jan 7;29(1):1-39.