Empirical Evaluation of Smart Phones Data Entry Using Four Different Keyboards

Nosaiba Dar Mousa
Systems Science and Industrial Engineering Department
State University of New York at Binghamton
Binghamton, USA
ndarmou1@binghamton.edu

Nabeel Mandahawi
Department of Logistics and Supply Chain Management,
Humber Institute of Technology and Advanced Learning,
Toronto, Canada
nabeel.mandahawi@humber.ca

Abstract

A study was performed to evaluate performance in entering English text on mobile devices by experienced users whose native language is not English and were experienced users of mobile devices and full keyboards. Two different mobile devices were used, they have virtual and physical keyboard, one of them have the full QWERTY keypad (single tap) and the other one have 14 keys keypad (multi-tap). Fifty subjects between the age of 19 and 44 years voluntary participated in the study, 22 of them are females and the rest are males. Subjects gave their opinions concerning learnability, accuracy, and efficiency. Subjects felt that the single tap method was easier to learn, more accurate and more efficient, than the multi-tap method, regardless of the type of keypad used.

Keywords
Virtual keypad, physical keypad and efficiency

1. Introduction

Sending electronic text messages, via mobile phones instead of a computer, is ubiquitous in industrialized societies. It is also fairly widespread in underdeveloped countries. The transmission of short textual messages (SMS texting) between mobile phones via the Wireless Application Protocol (WAP) to other systems and devices has been increasing exponentially, and integration with other services such as email and web services has also followed the same trend (Chang et al., 2017). Emailing, online chatting, and other web services all use text inputting. SMS texting is done mostly on mobile devices, such as personal digital assistants (PDAs), personal organizers, and cellular phones (O'Riordan et al. 2005). However, as O'Riordan et al. (2005) notes, many phone manufacturers target fashion trends and change mobile phone designs so as be different from their competitors, usually at the expense of user performance and usability. These changes and the impact of using a miniaturized keypad have introduced many factors that affect performance, usability and user satisfaction.

Many studies have been conducted to determine how the negative effects of these factors may be minimized or eliminated to gain greater user acceptance of mobile devices. The most common measures of usability of devices investigated are speed of text entry, accuracy of text entry, and subjective opinions of acceptability. The wide variety of factors that are thought to affect usability that have been studied include: (i) physical design features of the devices, such as type of keybpad, virtual or physical (Millet and Aydin 2010; Po-Hung Lin, 2015; Ruan et al. 2016), layout of the keys (O'Riordan et al. 2005; Lin 2017), mapping of letters on the keys, size of keyboard/keypad (Karlson et al, 2006; Lin 2017), and tactile feedback (Brewster et al. 2007; Dunlop and Taylor 2009; Hoggan et al. 2008 and Lopez et al. 2009); (ii) text input technology (Butts and Cockburn, 2002; James and Reischel, 2001; O'Riordan et al. 2005 and Wigdor and Balakrishnan, 2004); and (iii) user characteristics, such as age group (Millet and Aydin, 2010;), experience in using keypads and keyboards (Fleetwood et al. 2006; and MacKenzie and Zhang, 2001), gender (O'Riordan et al. 2005); and number of hands used (O'Riordan et al. 2005). The effects of these factors are not independent of one another. There are complex interactions among them, which often yield good information on performance and usability. The present study is focused subjective user opinions on the usability of mobile devices.

The most common method of entering text on mobile devices is direct character (letter) entry by tapping (pressing) on the keypad of the mobile device. Some devices have a physical (hard) keypad, others have a virtual (soft or touchscreen) one. The latter appears as an image on the device screen when the mobile device is activated for use. Unlike the normal sized computer keyboard, the miniaturized keypads on mobile phones and PDAs do not conform to the size and functional characteristics of the hands. The QWERTY keypad on mobile device has the keys for the 26 letters of the alphabet and other functions (e.g., for text navigation, text mode changes, and grammatical construction) packed within a small physical area. This leaves inadequate space for a person's finger to tap the desired key without touching an adjacent one (Mackenzie and Zhang, 2001; Sears et al. 1993). Therefore, the user is likely to slow down to minimize this problem. For example, Sears et al. (1993) demonstrated a drop-in speed of over 30% when using a small virtual keyboard compared to a large one that enabled the use of all fingers for keying.

Some mobile device keypads (mostly virtual ones) have a full QWERTY key arrangement, with each letter of the alphabet mapped onto a separate key. Others have two to four letters mapped onto a single key due to the limited area and number of keys on the keypad. The 12-15 keys on mobile phones, for example, have been adapted for this multi-mapping of letters. The second, third, or fourth letter is activated either by having the user tap several times (multi-tap) on that key to input a single letter, or rely on a mathematical method that uses a built-in dictionary to predict the word intended by the user, after tapping keys in sequence. Speed and accuracy are therefore hampered by the extra cognitive effort required for disambiguating the different letters on a specific key (Gong et al. 2005; James and Reischel, 2001; and Nesbat, 2003).

Different studies have been conducted to evaluate text entry and other performance characteristics on mobile devices are either empirical or based on mathematical theory. The mathematical studies have produced predictive models of text entry speed based on Fitt's law concerning target (key) selection, and on Hick-Hyman law concerning choice reaction time for visual scanning of the keyboard. However, criticism has been directed against possible misleading information suggested by the predictive models (Butts and Cockburn, 2002; James and Reischel, 2001). The models have been criticized for being too optimistic. Their predictions are based on a hypothetical 'perfect' user of a keypad, and suggest that ordinary users can attain the upper bounds of speed (54 wpm) that the models theorize (Fleetwood et al. 2006). It has, therefore, been suggested that further research in text entry should rely less on mathematical modeling and more on empirical testing, from the user's perspective, to influence mobile device design (Lopez et al. 2009; O'Riordan et al, 2005). The present study is an empirical one that examines user preferences (learnability, efficiency and accuracy) in text messaging from four different methods, using miniaturized physical and virtual keypads. No other single published study has examined these four methods. Moreover, this is the first such text entry study in the Middle East where a non-English speaking population widely uses imported mobile devices from English speaking countries.

2. Experimental design

The present study was conducted to study the significant of the factors on user preference for four text entry methods, from the perspective of ease of use, accuracy, and efficiency. These methods were single tap on a physical keypad, single tap on a virtual keypad, multitap on a physical keypad, and multitap on a virtual keypad. In addition, the evaluation was performed on a sample of subjects drawn from a population (Jordanians).

Subjects were asked to enter a set of selected text phrases as they would normally do, with respect to speed and accuracy. They were instructed not to go back (using the backspace key) to correct any recognized errors. The ability to produce error-free text was not an objective of the experiment. The four methods investigated in this study are the probably most popular among mobile device users in Jordan, not necessarily by choice, but because of the type of mobile devices available in the retail market. They are characterized by two different text input methods (single tap and multi tap), each on two types of keypads (physical and virtual). Gender was also an independent factor in the experiment. A repeated measure was used for gathering and analyzing the data, with each of the 4 factor combinations (text input method x keypad type). Subjects were nested within gender. The dependent variables were speed and accuracy of text entry, measured in words per minute and number of errors per character, respectively. The repeated measures assignment minimized the impact of the variability (in performance measures) among subjects. Two different mobile devices were used in the experiment, each with a virtual and physical keypad. The first one is physical and virtual with full QWERTY keypad and the second one is virtual and a condensed physical keypad, each with 14 keys. A custom designed software application was used in each device to record the time for text entry accurately, and to store the entered text for subsequent error counting. After the text entry tests, subjects were asked to complete a short questionnaire to subjectively evaluate the devices and entry methods. The questions (Butts and Cockburn, 2002) related to learnability (I find this method easy to earn), accuracy (I did not make any mistakes), and efficiency (This input method is efficient to use).

3. Subjects

Fifty subjects, 22 females and 28 males, between the ages of 19 to 44 years, (mean=23 years) volunteered to participate in the study. Twenty two were university students (6 from medicine, 3 from pharmacy, and 13 from engineering), and 28 were employees in IT companies (programmers and engineers). The participants were selected because of their experience in using mobile devices and keyboards/keypads. They were all self-declared heavy users of mobile devices and computers. The effect of different levels of experience was not under investigation so we decided to use subjects with a uniform level of experience. There was no restriction on hand dominance. The age of subjects was also not controlled within any range. Five subjects were left handed and 45 right-handed. All were native Arabic speakers who use English as a second language in Jordan.

4. Text Phrases Selection

Five English language phrases, shown below were used for text entry in the experiment. They were taken from James and Reischel (2001) to allow for comparison of results with these studies. The five phrases were of moderate word lengths, and were also selected because they did not include punctuation or syntax. The experiment was designed to eliminate the effect of grammar, which could have affected comprehension and, hence, speed and error in text entry. The phrases and the number of characters (in parentheses) are as follows:

- what show do you want to see (28)
- let me know if we should wait (29)
- hi joe how are you want to meet tonight (39)
- want to go to the movies with sue and me (40)
- we are meeting in front of the theater at eight (47)

5. Results and Discussion

The Friedman non-parametric ANOVA was used to analyze the ranked data, for each of the three questions, from the questionnaire, using the score on the 5-point Likert scale (1=strongly disagree and 5=strongly agree) as the dependent variable and method of text entry (with four levels) as the independent variable (Table 2). The results indicated that subjects did not agree that all four methods of text entry had the same level of learnability (Friedman statistic, S= 52.58; p= 0.000), error rate (S= 15.29; p= 0.002), or efficiency (S= 55.89; p= 0.000). The Friedman's test was followed by Tukey's multiple comparison test to determine which of the four methods differed from the others, as displayed in Table 2.

Table 2. Mean (standard deviation) responses to subjective questions, on the 5-point Likert scale

	Q1. I found this method easy to learn (learnability effect)	Q2. I did not make many mistakes with this method (accuracy effect)	Q3. Overall, this input method was efficient to use (efficiency effect)
Virtual + single tap	4.16 (0.87)	3.60 (1.33)	3.96 (0.67)
Physical+ single tap	4.26 (0.80)	3.72 (1.23)	4.24 (0.72)
Virtual + multitap	2.86 (1.29)	2.52 (0.93)	2.68 (0.99)
Physical multitap	3.34 (1.00)	2.88 (0.96)	3.16 (0.93)
Single tap average for virtual and physical	4.21	3.66	4.1
Multitap average for virtual and physical	3.1	2.7	2.92
Physical average for single and multitap	3.8	3.3	3.7
Virtual average for single tap and multitap	3.51	3.06	3.32

For each column (learnability, accuracy or efficiency), the scores next to a common line are not significantly different (p>0.05), otherwise they are significantly different from one another. The results indicate that learnability, accuracy and efficiency were all significantly better with single tap text entry than multitap, regardless of the type of keypad used (average score = 4.21 vs 3.10, 3.66 vs 2.70, and 4.10 vs 2.92 for learnability, accuracy and efficiency, respectively); and the physical keypad was rated slightly better than the virtual one for learnability, and accuracy, but the difference, in each of the two cases, was not significant, while the efficiency was significantly better with physical keyboard than virtual keyboard (3.7 vs. 3.32).

Even though the speed from the physical keypad was faster (10.08%) than from the virtual keypad, subjects did not feel that the physical keypad was easier to learn or more accuracy. This apparent discrepancy can be explained by the fact that subjects were already expert users and had already reached the point where learning did not increase significantly with time or use. It is likely novices may find that text entry on the physical keypad is easier to learn than on the virtual keypad. It is also not clear why subjects did not find accuracy with the physical keypad significantly greater than with the virtual keypad, for single tapping or multitapping. It is surprising that, even though subjects made significantly less errors with the physical keypad, they did not think they were more accurate with it compared with the virtual keypad. Again, this may be due to familiarity with both types of entry method. The small differences from objective measurements (1.93 vs 2.98 in Table 1) that are statistically significant may not be practically significant in the minds of subjects.

At the end of the experiment, subject's general thoughts concerning the PDA design were collected. Some subjects didn’t feel comfortable with all the input methods; for example, eight participant's thought that there was a lag in time between the key press and the appearance of the letter on the virtual (touch) screens for the multi press input. Twenty eight subjects declared that their multi press performance could be enhanced with more frequent use of this method. Furthermore, due to the difference between the QWERTY multitap keyboard layout and the ABC layout, which is used on most of the Nokia devices, subjects reported that they have to focus at the keypad whenever they have to press the next key rather than focusing on the screen, which they prefer. All subjects, especially the IT ones and the university students, stated that the chat abbreviations that they normally use would have been easier to use for sending text messages.

Hand dominance has not been examined in this study, but there is no evidence in the literature that the methods of text entry tested in this study are affected by whether subjects are left or right handed. We also did not test for differences due to number of hands used in the study, but recent evidence (Millet and Aydin, 2010) indicates that the use of two hands (two thumbs) to enter text was not significantly faster than the use of one hand (only 5.7% faster). It’s possible that the use of the stylus may not have made a difference in the results of the present study since the duration of text input was relatively short.

6. Conclusions

This study supports the generally held view that text entry speeds and error rate depends on the type of keypad used and the method of entering the characters. It also shows that this view is not simplistic. In general, subjects didn't feel comfortable with all the input methods, eight thought that there was a lag in time between the key press and the appearance of the letter on the virtual (touch) screens for the multi press input. Twenty eight subjects declared that their multi press performance could be enhanced with more frequent use of this method. Further research paper will focus on the objective analysis to study the correlation between the speed of text entry and the error rate.

References

Brewster, S., Chohan, F., Brown, L., Tactile feedback for mobile interactions, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, USA, April 28-May 03, 2007.

Butts, L., and Cockburn, A., An evaluation of mobile phone text input methods, *Australian Computer Science Communications*, vol. 24, no. 4, pp. 55 – 59, 2002.

Chang, J., Choi, B., Tjolleng, A., and Jung, K., Effects of Button Position on a Soft Keyboard: Muscle Activity, Touch Time, and Discomfort in Two-Thumb Text Entry, *Applied Ergonomics, vol.* 60, pp. 282–292, 2017.

Dunlop, M.D., and Taylor, F., 2009. Tactile feedback for predictive text entry, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, USA, April, 04-09, 2009.

Fleetwood, M., Byrne, M., Centgraf, P., Dudziak, K., Lin, B., and Mogilev, D., An evaluation of text entry in Palm OS - Graffiti and the virtual keyboard, *ACTA Press, Anaheim, CA, USA*, 2006.

Hoggan, E., Brewster, S.A., and Johnston, J., Investigating the effectiveness of tactile feedback for mobile touch screens, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, USA, April 05-10, 2008.

James, C.L., and Reischel, K.M., Text input for mobile devices: comparing model prediction to actual performance, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* USA, March 31-April 05, 2001.

Karlson, A., Bederson, B., and Contreras-Vidal, J., Understanding single-handed mobile device interaction, *Proceedings of the Canadian Information Processing Society*, Canada, 2006.

Lin, P.,., Investigation of Chinese text entry performance for mobile display interfaces, *Ergonomics*, vol. 58, no. 1, pp. 107-117, 2017.

Lopez, M.H., Castelluci, S., and MacKenzie, I.S., Text entry with Apple iPhone and the Nintendo Wii. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* USA, 2009.

Lopez, M.H., Castelluci, S. MacKenzie, I.S., Text entry with Apple iPhone and the Nintendo Wii. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, USA, April, 04-09, 2009.

MacKenzie, I. S., and Zhang, S.X., An empirical investigation of the novice experience with soft keyboards, *Behaviour and Information Technology*. Vol. 20, no. 6, pp. 411-418, 2001.

Millet, B., and Aydin, B., Empirical evaluation of text entry performance of the Apple iPhone and a hard key mini-QWERTY keyboard smartphone, *Proceedings of the XXIInd Annual International Occupational Ergonomics and Safety Conference,* USA, 2010.

Nesbat, S. B., A system for fast, full-text entry for small electronic devices, In *Proceedings of the 5th international conference on Multimodal interfaces*, Canada, Nov 5-7, 2003.

O'Riordan, B., Curran, K., and Woods, D., Investigating text input methods for mobile phone, *Journal of Computer Science*, vol. 1, no. 2, pp. 189-199, 2005.

Ruan, S.,Wobbrock, J.O., Liou, K., Ng, A., and Landay, J., Speech is 3x faster than typing for English and mandarin text entry on mobile devices, *arXiv preprint arXiv:1608.07323*, 2016.

Sears, A., Revis, D., Swatski, J., Crittenden, R., and Shneiderman. B., Investigating touchscreen typing: the effect of keyboard size on typing speed. *Behaviour and Information Technology*, vol.12, no. 1, pp.17-22, 1993.

Wigdor, D., and Balakrishnan, R., A comparison of consecutive and concurrent input txt entry techniques for mobile phones, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Austria, April 24-29, 2004.

Nosaiba Dar Mousa is a PhD student in Systems Science and Industrial from Binghamton University with more than 17 years of professional experience in different industries. She has a B.Sc in Mathematics from University of Jordan (2002), and M.Sc in in Industrial Engineering from University of Jordan (2009). Her research interests

include data and decision science, and supply chains. Nosaiba is a member at Alpha Pi Mu, and Who's Who Award.

Nabeel Mandahawi is an Associate Professor at Humber Institute of Technology and Advance Learning, Toronto, Canada. Mr. Mandahawi completed his Ph.D. in Industrial Engineering from the Department of Industrial, Manufacturing, & Systems Engineering at University of Texas at Arlington. (2004), M.Sc. in Industrial Engineering from University of Jordan (2000), and a B.Sc. in Mechanical Engineering from Jordan University of Science and Technology (1997). Prior to joining Humber, Mr. Mandahawi worked for more than eight years as a Professor at both the graduate and the undergraduate levels in Jordan. Furthermore, he has extensive industrial experience since 1997 in Operations and Supply Chain Management as a mechanical engineer, consultant, trainer, and quality assurance engineer for diverse range of manufacturing, healthcare and service sectors. Furthermore, he has two years of professional educational planning experience as lead for accreditation of the engineering degrees by the Accreditation Board for Engineering and Technology and an EFQM assessor promoted by the European Foundation for Quality Management where he worked as an excellence evaluator for different international excellence programs. His research interests include operations management, logistic and supply chain management, simulation modeling and ergonomic. He is a member at Tau Beta Pi, Alpha Pi Mu, and Who's Who Award.