A Qwirky Kind of Qwerty: A study of a Qwerty keyboard with *t* and *f* swapped

Blaise Calaycay
University of Toronto
Toronto, Canada
blaise.calaycay@mail.utoronto.ca

Derek Choi
University of Toronto
Toronto, Canada
derek.choi@mail.utoronto.ca

Kelley Tai
University of Toronto
Toronto, Canada
kelley.tai@mail.utoronto.ca

ABSTRACT

Our Qwirky keyboard is a soft Qwerty keyboard with the letters *t* and *f* swapped. *T* and *f* are swapped based on English word and digraph frequency. This is intended to create a more efficient keyboard while minimizing learning time. We focused on soft keyboards because hard keyboards have been around longer than soft keyboards. In a controlled experiment, we compared our design with the Qwerty keyboard in a series of sessions. 14 participants volunteered over the course of 7 sessions for 4 - 7 days. Users started the Qwirky keyboard with an average of 18 words per minute (WPM) and an accuracy rate of 82% and peaked at 32 WPM and 91% accuracy on or before the fifth session. However, participants experienced a decline for both categories after the fifth session, making the data inconclusive and necessitating further study.

Author Keywords

Soft keyboards, text entry, mobile technology, digraphs, linguistic model, words-per-minute (WPM), t-transition, f-transition, t-difference, f-difference

ACM Classification Keywords

H.5.2 [User Interfaces]: Input devices and strategies

INTRODUCTION

The Qwerty keyboard has become a standard keyboard layout since its introduction in the 19th century. However "universal" the keyboard is for alphabetic writing systems, it is, arguably, not as efficient as it could be. There have been many alternatives keyboard designs to Qwerty, but most users have mastered the Qwerty keyboard. Because of this, the motivation to learn a "more" intuitive keyboard is very low. Switching from the Qwerty keyboard to another keyboard, such as Dvorak or Opti [6], is such a large commitment that most users may not be willing to learn it. If those non-Qwerty keyboards cannot be adaptable because they are too complex to learn, and Qwerty is arguably inefficient, what is the best way to create the most optimal keyboard?

Origins of Soft Keyboards

There have been many Human Computer Interaction studies on keyboard optimization. The Qwerty keyboard has been around since the 19th century. This has given researchers more than 200 years of tinkering with hard keyboard layouts whereas soft keyboard research has only been around for the past 25 years [5]. However, one can argue that soft keyboards did not reach its mainstream status until 2007, when Apple started selling the original iPhone. This means smartphone users have been using the Qwerty layout for more than eight years. Relatively speaking, typing on soft keyboards is new compared to typing on personal computers, but eight years is plenty of time for soft keyboard users to be trained on the traditional layout.

Although the Qwerty keyboard has become the standard for almost all modern devices, it does not suggest that it has surpassed all alternatives in terms of efficiency. Bi et. al's study shows that multi-character key sets produce more characters per second than the traditional Qwerty keyboard [2]. In one of MacKenzie et. al's soft keyboard study, the researchers created a layout that was more efficient than Qwerty [6]. However, to increase the likelihood of the general public adopting a new layout, efficiency alone may not be enough. The time for a user to learn and become familiar with the new layout must be minimal. When creating a new layout, the researchers must consider that different layouts have different learning curves [1].

Experiments on Soft Keyboards

In Anderson et. al's study [1], the average task completion time for the Qwerty keyboard was 40 seconds. The average times for the fifth trial on the chord, contoured split, Dvorak, and split fixed-angle keyboards were 346 seconds, 69 seconds, 181 seconds, and 42 seconds, respectively [1]. In addition, learning percentages were calculated based on user ratings to quantify the learning demands of each keyboard. Learning percentages for the chord, contoured split, Dvorak, and split fixed-angle keyboards were 77.3%, 76.9%, 79.1% and 90.4%, respectively [1].

These results suggest that depending on the layout and ergonomic factors with each alternative keyboard, the learning curve will vary greatly. It is important to keep in mind that some of these keyboards mainly required physical learning whereas others required mostly cognitive learning, and some that required a combination of both. Thus, the speed at which one becomes proficient at an alternative keyboard will also vary from person to person depending on their physical or cognitive adaptability.

In the second study [5], participants reached a faster typing rate (in WPM) on the OPTI keyboard than on the Qwerty keyboard in only 20 sessions on average (equivalent to 4 hours). This shows that efficient alternative layouts, if created correctly, can take relatively minimal time to learn. It is important to note that since people are more well-versed on soft keyboards now than at the time of the study, the general population's learning curve for the OPTI layout may have increased since then. Therefore, it was important for us to find a compromise between learning time and efficiency. Xiaojun, Smith, and Zhai [5] show how this could be achieved by minimizing the changes from the traditional Qwerty layout. In the study [5], the Quasi-Qwerty layout had a few number of keys shifted to one key-distance from their original position. There is plenty of room for improvement on this study. Modifications to this study, such as not limiting a key to only move a key-distance, or swapping only two to three keys, could prove to be insightful.

QUESTION AND HYPOTHESIS

Will changing the placement of two keys on the Qwerty keyboard result in a more accurate and/or faster typing experience? Based on data collected from multiple sessions (e.g. WPM, learning curve), the study compares the differences in performance between the two keyboards.

T and *f* are swapped for multiple reasons. "Th" is the most common bigram in the English language [2]. Having both *t* and *h* in the home row, with *t* mapped to the left hand and *h* mapped to the right hand, will hopefully speed up typing the combination. Of the 50 most frequent words in the English language, the words that are composed of the letter *f* are also exclusively made up of letters on the top row, with the exception of the word "from" [2]. In addition, with our placement of *f*, these common words can be typed by alternating hands for each letter, which would hopefully result in faster typing. Furthermore, *t* and *f* are only one key away from each other, thereby minimizing the layout change. We considered not to place *t* and *h* next to each other because we wanted to avoid possible finger collisions.

We predicted that although we would see an increasing learning curve in the Qwirky keyboard, the Qwerty keyboard would be too powerful to be beaten. Since most users are too comfortable using the Qwerty keyboard, our final results might not be significant.

METHODOLOGY

In this section, we describe the methodology that helped us address our research question and hypothesis.

Apparatus

An Android mobile phone application (app) called Qwirky was created for this study. This app (minimum SDK: API 21 Android 5.0 Lollipop) was installed on an LG Nexus 5 with a 2GB Snapdragon 800 RAM, a 4.95 inch display (70.8% screen to body ratio), and a 1080x1920 pixel screen resolution.

The layout of the keys in the modified keyboard included in Qwirky is identical to the Qwerty keyboard except for keys *t* and *f*. In our keyboard, *t* and *f* are swapped (Figure 1). Each key has width equal to 10 percent of the screen width, height equal to 60dp (density-independent pixels) and is 0 pixel apart from each other (although aesthetically it looks like it has a tiny gap).

Included in the app is a Qwerty keyboard similar in appearance to our keyboard, with the only difference being *t* and *f* are swapped.

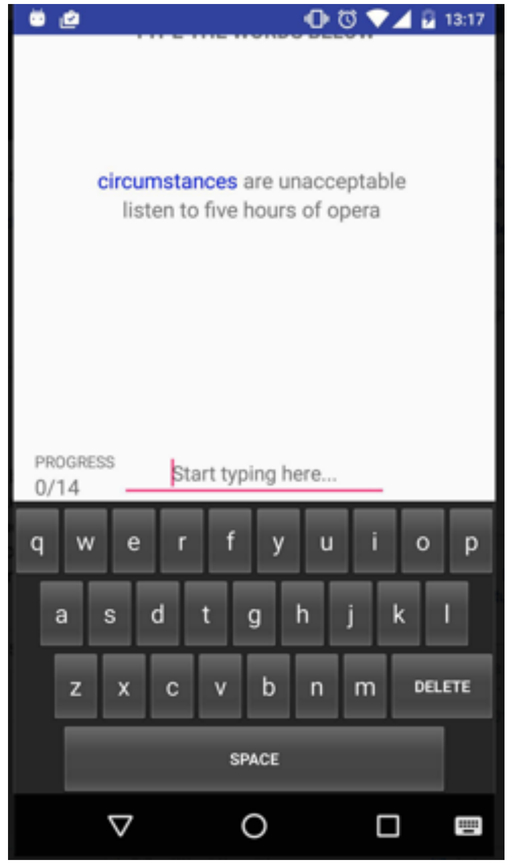

Figure 1: Qwirky app

In the app, the user is shown a phrase they will have to type (Figure 1). There is a blank text field where the participant types the words they see on the screen. Typing is done on a word-by-word basis. This means that when users finish typing a word (by pressing the spacebar or an appropriate punctuation mark), the word disappears from the text field, which is left blank for the next word.

The app recorded each session's WPM, accuracy, number of *t*'s typed, total *t-transition* time, number of *f*'s typed, total *f-transition* time (transition times discussed in Data Collection).

Participants

14 participants volunteered to be part of our study, but only 12 completed the study in its entirety. These participants all so happened to be 20 - 24 year old students studying either computer engineering or computer science. It also just so

happened that 4 of these participants were female and the other 10 were males. Although it was a priority to find participants who were avid Android users, only 11/14 participants were Android users. All Android participants used the default Google keyboard as their main keyboard. All participants were comfortable typing on soft keyboards, and all the participants most preferred method of inputting text was with two fingers on both hands.

Experimental design

This experiment was a within-subject study with two factors: the keyboard layout (Qwirky and Qwerty) and the number of sessions (7 sessions). Therefore, our independent variable has two levels: The Qwirky keyboard and the Qwerty keyboard. We compared these two levels to see which keyboard was more efficient so it was only natural for our study to be within-subject.

Each session lasted about 10 minutes and each session separated into two 4-minute half-sessions for each keyboard layout. There was a 1-2 minute break between the half-sessions. Each half-session contained 14 phrases. These phrases were chosen from Norvig's [1] modernized study of Mark Mayzner's corpus. Here is an example of a phrase we will make our participants test out. Again, this is taken from [6].

Each session contained the same phrases for both keyboard layouts. Therefore, 98 phrases were chosen at random.

Each participant completed 7 sessions. These sessions were scheduled whenever the participants were free. Each session was separated by at least 2 hours and at most 2 days. This was to ensure that the participants were learning. We also didn't wish for the participants to feel rushed and urged to complete the task and encounter fatigue.

There are different text entry methods on soft keyboards. It was important for each participant was required to type on the keyboards with two fingers on both hands. This was the preferred method of inputting text for all of participants too. The orders of the two keyboards that participants had to type were different. Half the participants started with the Qwerty keyboard and half of the participants started with the Qwirky keyboard.

Tasks and procedures

Before the experiment began, each participant was given a welcoming session. This welcoming session consisted of a debrief on the purpose of our study. Participants then filled out consent forms. We emphasized that participants would not get any reward (money and gifts) for being in our study and that they could stop at any time they wanted. These volunteers were also given a pre-experiment questionnaire which consisted of questions about a person's background on using soft keyboards. Before the first session began, participants was given a trial session. This trial session contained 7 phrases that were randomly chosen from Mark Mayzner's corpus as well. This took around 1 minute to complete.

All the experiments were conducted in a quiet environment with no interruption. For example, a cubicle in the workplace or a study room in the university. All participants sat down while typing on the both keyboards.

Each session contained a set of phrases from a large corpus. Each participant was required to type those phrases on both keyboards. Half of the participants started on the Qwerty keyboard and half the participants started on the Qwirky keyboard. The same set of phrases was used in each session. For each session would have its own list of phrases from Mark Mayzern's corpus [6]. This was to eliminate potential bias that users might remember the phrases they typed before. After the experiments were done typing on one of the keyboard layouts, there was a brief break before they moved onto the other keyboard.

Measures

Our two dependent variables were the WPM and accuracy. WPM was one of the most basic determinants for the efficiency of a keyboard. Since our objective was to design a more optimal keyboard, it was important to compare the WPM between Qwirky and Qwerty. We wanted to see how fast the participants can type on both keyboards because a supposedly "good" keyboard yields higher words per minute.

The other dependent variable is accuracy. Good accuracy means fewer keystrokes, and fewer keystrokes allow users to achieve typing goals faster. Accuracy also implies that there are fewer misspellings in words. We wanted to see if our keyboard had all those qualities, so it was important to compare WPM and accuracy to both of these keyboards.

Data collection

When the app is run and a participant starts typing, the app collects the necessary data in the background. This includes WPM, accuracy, number of *t*'s typed, total t-transition time, number of *f*'s typed, total f-transition time.

WPM is measured as Net WPM. The formula for Net WPM is as follows: [(number of typed entries / 5) - uncorrected errors] / time in minutes. We only counted uncorrected errors since correcting mistakes already penalizes the user by using up time. We also divide the total number of entries by five, as five is the average length of each word. Percentage of accuracy is calculated by dividing the number of correct characters typed by the total number of characters typed. T-transition times are the total time that the user transitioned from a character to *t* and transitioned from *t* to a different character. This definition applies to f-transition time, but on *f*. The average transition times per instance of *t* and *f* were compared between Qwerty and Qwirky to determine if there was any effect on the speed that it takes the user to type *t* and *f*.

Finally, at the beginning of the experiment and at the end of the last session, we collected information from each participant regarding their impression of the Qwirky keyboard. Information was collected in NASA Task Load

Index (NASA TLX) format. Questions regarding the workload of our experiment were presented with a grading scale that participants filled out based on their experience.

After each session, data was sent to the participant's email. The data was then fed to a Python script to calculate the average WPM for each session per keyboard, average accuracy for each session per keyboard, and other interpretations necessary. We also fed the data on Mackenzie's ANOVA program [6] to calculate repeated-measures ANOVA and other calculations of analysis of variance.

RESULTS

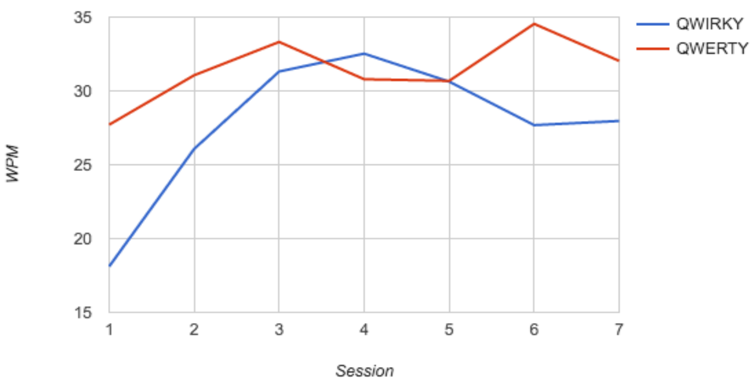

Figure 2: Average words-per-minute over 7 sessions

A plot of the learning curve shows that for Sessions 1 - 4, the average WPM of the 12 participants continually improved. Furthermore, the first instance that the average WPM for Qwirky exceeded that of Qwerty was on Session 4. However the WPM for Qwirky decreased from Sessions 5 - 7, hovering below the WPM for Qwerty. Dives in WPM can be seen in longitudinal keyboard studies like Mackenzie's [6], so this does not necessarily mean that Qwirky is permanently worse than Qwerty in WPM. Unfortunately, an analysis of variance on WPM showed no significant effect for keyboard ($F_{1,11}$ = 0.153, $p > .05$), session ($F_{6,66} = 1.7$, $p > .05$), or keyboard-by-session interaction ($F_{6,66} = 1.88$, $p > .05$). This could be due to multiple reasons, such as too few sessions.

On the other hand, an analysis of variance on accuracy showed a significant effect for session ($F_{6,66} = 3.06$, $p < .011$) (no significant effect for keyboard and keyboard-session interaction). Digging further showed that accuracy for Qwerty was not significant for session ($F_{6,66}$ = 1.22, $p > .05$), but the accuracy for Qwirky was notably significant for session ($F_{6,66} = 4.46$, $p < .00081$). Furthermore, an analysis of variance on accuracy based on the first keyboard showed significant effects for session on Qwirky accuracy for participants whose first keyboard was Qwirky ($F_{6,30} = 4.038$, $p < .0045$) but not for participants whose first keyboard was Qwerty ($F_{6,30} = 1.798$, $p > .05$).

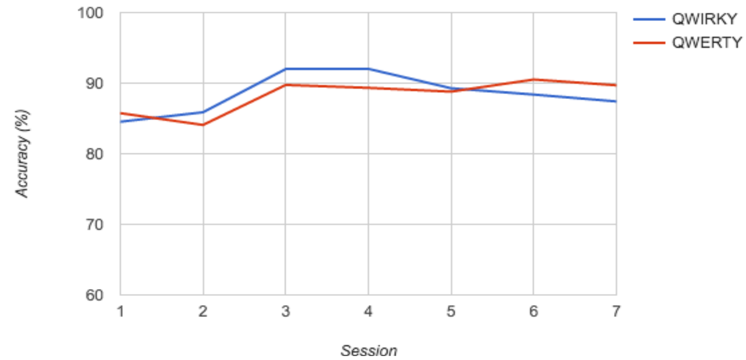

Figure 3: Average accuracy over 7 sessions

Since sessions 5 - 7 showed a decrease for Qwirky on both WPM and accuracy, we analyzed the difference between the number of *t*'s typed and the number of *t*'s in the phrase set for each session (we call this *t-difference*). The data shows that from Sessions 1 - 5, the average t-difference continually decreased. This means continual improvement was seen when using Qwirky. Surprisingly however, participants showed a significant regression for Sessions 6 and 7. The t-difference was 6.4 for Session 6 and 5.25 for Session 7. F-difference showed similar behavior, although *t* occurs significantly more in the phrase sets than *f* (average of 7 vs 32).

Session	QWERTY	QWIRKY
1	3.083333333	7.916666667
2	1.833333333	3.5
3	0.3333333333	2.833333333
4	3	2.583333333
5	0.75	1.291666667
6	2.583333333	6.416666667
7	2.708333333	5.25

Figure 4: Average t-difference per session

Average F- and t-transitions were similar for both keyboards.

DISCUSSION

WPM and accuracy drops

We saw an increase in accuracy and WPM for the Qwirky keyboard during the first four sessions. For accuracy, Qwirky "caught up" with Qwerty by the second session. For WPM, Qwirky caught up by the fourth session. Although the fourth session for Qwerty saw a dip in the WPM curve, dips like this are normally seen in similar studies [6]. Therefore, Qwirky catching up could be seen as a good achievement. However, both accuracy and WPM experienced declines after the fourth sessions.

It was difficult to determine why both categories experienced drops in their learning curves. For accuracy, t-difference was a significant factor in the decline. T-difference for Session 5 increased by almost 400% – a complete u-turn from the pattern seen in the first four sessions, where improvement rates of 50% were seen.

A reason for these declines could be fatigue. In post-experiment interviews, four participants mentioned that towards the end of the 7-session experiment, they became indifferent towards the study. One participant said she "stopped caring about the experiment" near the end of the study. The effect of indifference showed not only in t-differences for Qwirky, but for Qwerty as well. Data of t-differnce for Qwerty showed a pattern similar to Qwirky (except for Session 4) in that the average t-difference increased in Sessions 6 - 7. Despite the breaks between keyboards, 14 phrases per keyboard may have been too tedious for each participant. Data from the post-experiment questionnaire showed that the participants thought that the mental demand for the experiment was high.

The keyboard layout itself could have affected accuracy. Most of the participants complained of either hitting a letter key when they intended to hit the spacebar (or vice-versa), or hitting a letter key when they intended to hit the delete key (or vice versa). The gap between the delete key and its adjacent keys on the default Google keyboard is aesthetically much wider than the same gaps on our keyboards. This is the same with the spacebar and its adjacent keys. A better keyboard layout with bigger key gaps could have yielded better data.

Significance of Qwirky accuracy on Qwirky-first

The data shows that participants who started each session with Qwirky showed an accuracy learning curve that was statistically significant. However, it was not statistically significant for participants who started with Qwerty. We hypothesize that this is due to priming. Priming is a when an exposure to a stimulus effects a response to another stimulus [3]. We believe that if participants started with the Qwerty keyboard as well, hey will think be primed with the Qwerty keyboard since they are so comfortable using the Qwerty keyboard. Therefore, when they start typing with Qwirky keyboard in the next half of the session, they will think that they are using the past keyboard since it so familiar. However, if participants started with Qwirky, the priming effect has no use. (This could potentially be because the Qwirky keyboard is too foreign). Therefore, the results were more significant for users who start with the Qwirky keyboard.

CONCLUSION

We have described the design and performance of our Qwirky keyboard for mobile devices. Our results showed that although participants showed an increase in WPM and accuracy for the first four sessions, both categories decreased starting from the fifth session. This may be due to fatigue and an attitude of indifference in performance towards the end of the study. Thus, the data is inconclusive. Therefore, this study finds that the null hypothesis is true: Qwirky is less efficient and accurate when compared to Qwerty.

Our biggest limitation was not having enough sessions. Although this is a unique point because it seemed that the typing on the Qwirky keyboard was mentally challenging. We also found that both keyboards we used was a difficult keyboard layout to adapt to.

Further experiments (ideally with more sessions) and a better keyboard layout are necessary to improve the study.

ACKNOWLEDGMENTS

Thank you for all the participants for helping us completes this experiment. We would also like to thank Ming Ming for being encouraging and giving us sound advice.

REFERENCES

1. Allison M. Anderson, Gary A. Mirka, Sharon M. B. Joines, David B. Kaber. 2009. Analysis Of Alternative Keyboards Using Learning Curves. Retrieved Feb 2, 2016 from http://hfs.sagepub.com/content/51/1/35.short

2. Xiaojun Bi, Barton A. Smith, Shumin Zhai. 2014. Quasi-Qwerty Soft Keyboard Optimization. Retrieved Feb 2, 2016 from http://dl.acm.org.myaccess.library.utoronto.ca/citation.cfm?id=1753367&CFID=749110248&CFTOKEN=21122957

3. Raluca Budiu. Priming and User Interfaces. 2016. Retrieved March 10, 2016 from: https://www.nngroup.com/articles/priming/

4. Darryl Francis. 2015. DVORAK and COLEMAK Keyboards. Retrieved March 10, 2016 from http://digitalcommons.butler.edu/wordways/vol48/iss4/17

5. John G. Kreifeldt, S. L. Levine, C. Iyengar. 1989. Reduced Keyboard Designs Using Disambiguation. Retrieved Feb 2, 2016 from http://pro.sagepub.com/content/33/6/441.short

6. I. Scott MacKenzie, Shawn X. Zhang. 1999. The design and evaluation of a high-performance soft keyboard. Retrieved Feb 2, 2016 from

7. Scott MacKenzie. Class Anova2. Retrieved March 10, 2016 from: http://www.yorku.ca/mack/Anova2.html

http://dl.acm.org.myaccess.library.utoronto.ca/citation.cfm?id=302983&CFID=749110248&CFTOKEN=21122957

8. Peter Norvig. 2012. English Letter Frequency Counts: Mayzner Revisited or ETAOIN SRHLDCU. Retrieved February 23, 2016 from http://norvig.com/mayzner.html

9. Amanda L. Smith, Barbara S. Chaparro. 2015. Smartphone Text Input Method Performance, Usability, and Preference With Younger and Older Adult. Retrieved March 10, 2016 from http://hfs.sagepub.com/content/57/6/1015