Thorough Software Layout Machine Analysis

Kristian Tichota

March 10, 2021

0.1 Abstract

This analysis investigates some claims about typing ergonomics using three large corpora of varying language complexities and SteveP's fork of the Patorjk analyzer. It ranks the software layouts by the scores assigned by the analyzer. It concludes that for the layouts tested, variations in language complexity have no significant impact on typing ergonomics on Ergodox keyboards while a higher language complexity results in significantly higher scores on ANSI keyboards. Even when punctuation is present, all of the tested alternative layouts score significantly higher than QWERTY. All layouts tested score significantly higher on Ergodox than ANSI keyboards. According to the test method, Colemak-DH and MTGAP score the highest on ANSI and Ergodox keyboards.

0.2 Introduction

There have been many attempts to create alternative software layouts to improve upon QWERTY, the two most notable of which are Dvorak [1] and Colemak [2]. The majority of popular layouts are based on Colemak and/or Dvorak. Despite the growing number of claims regarding typing ergonomics, such as Workman's claim that same finger utilization does not influence ergonomics [3], or that ergonomic keyboards do not influence ergonomics [4], there has not been enough scientific research [5] and the bulk of machine analysis has not been sufficiently thorough. The goal of this analysis is to thoroughly investigate some of those claims while avoiding pitfalls such as inadequate sample size, inaccurate analyzer, poor interpretation of results, and bias.

0.2.1

Primary claims to be investigated:

1. Alternative layouts are only good for common English
2. Alternative layouts fail when punctuation is present
3. Layout "x" is better/worse than "y"
4. Ergonomic keyboards do not influence ergonomics significantly

0.2.2

Layouts included in this analysis, in ANSI and Ergodox variants:

- QWERTY
- CarpalX-QGMLWY
- Colemak
- Colemak-DH
- Dvorak
- Dvorak-ID
- Halmak
- Capewell
- Norman
- Workman
- Asset
- MTGAP

All ANSI layouts in this anlalysis use a straight left wrist fingering scheme, also known as the angle principle [6] [7], for the left half bottom row. In the traditional typing technique [8], the left hand is twisted for the bottom row keys, causing harmful ulnar deviation strain [9]. This is due to the fact that hands are mirror symmetrical, while the standard row-staggered keyboard is not.

0.3 Method

The layout analyzer used for this analysis is SteveP's fork of the Patorjk analyzer [10]. Its scoring is based on distance calculation (5/10 weight), finger distribution (2/10 weight), and same finger bigrams (3/10 weight). The other two popular analyzers are CarpalX [11] and Patorjk's analyzer [12]. Every mentioned analyzer has issues; however, CarpalX and Patorjk's analyzer use the traditional typing technique as their basis and have inaccuracies in raw calculations while SteveP's analyzer only omits one penalty and is based on the angle principle, making it ideal for this analysis.

0.3.1

Issues with the analyzers:

- CarpalX has consistent but inadequate base effort chart based on the traditional typing technique [13].
- Patorjk's analyzer weighs in hand alternation with no respect to comfortable rolls, such as ST on Colemak. Its distance calculation is inaccurate because of not taking wrist angles into consideration and not distinguishing the more difficult lateral movement from vertical. Its finger distribution heavily favours the middle finger, which promotes tendon overload. It is also based on the traditional typing technique [14].
- SteveP's analyzer does not take prolonged consecutive use of the same hand into consideration.

0.3.2

Three corpora of different complexities were used for this analysis:

- Complex English
- Word Count: 1,675,921
- Flesch Reading Ease Formula: 52.46 (12.81 Flesch-Kincaid)
- FOG Index: 15.06
- SMOG Index: 12.62
- Automated Readability Index: 13.88

- Plain English
- Word Count: 1,869,673
- Flesch Reading Ease Formula: 71.95 (7.56 Flesch-Kincaid)
- FOG Index: 9.71
- SMOG Index: 9.36
- Automated Readability Index: 7.10

- Easy English
- Word Count: 1,732,871
- Flesch Reading Ease Formula: 78.09 (6.68 Flesch-Kincaid)
- FOG Index: 8.66
- SMOG Index: 8.07
- Automated Readability Index: 6.35

0.3.3

The total volume of the three corpora is 5,278,465 words. All three corpora contain books from Project Gutenberg while the complex English corpus also contains the entirety of Lovecraft's fictional work.

- Books used for complex English:
- The entirety of Lovecraft's fictional work
- Odyssey and Illiad by Homer
- The World as Will and Idea (Vol. 1-3) by Arthur Schopenhauer

- Books used for plain English:
- The Brothers Karamazov by Fyodor Dostoyevsky
- Holmes' Own Story by Herman W. Mudgett
- Some do not... by Ford Madox Ford
- Dracula by Bram Stoker
- Other Worlds by Lena Jane Fry
- Moby Dick; Or, The Whale by Herman Melville
- The Hound of the Baskervilles by Arthur Conan Doyle
- City of Comrades by Basil King
- Dubliners by James Joyce
- The Republic by Plato
- Bleak House by Charles Dickens

- Books used for easy English:
- Peter Pan by J. M. Barrie
- The Wonderful Wizard of Oz by L. Frank Baum
- Anne of Avonlea by L. M. Montgomery
- Kim by Rudyard Kipling
- The Blue Fairy Book by Andrew Lang
- The Princess and the Goblin by George MacDonald
- At the Back of the North Wind by George MacDonald
- Treasure Island by Robert Louis Stevenson
- Little Women by Louisa May Alcott
- The Secret Garden by Frances Hodgson Burnett
- Anne of Green Gables by L. M. Montgomery
- Five Children and It by E. Nesbit
- The Children of the New Forest by Frederick Marryat
- The Orange Fairy Book by Andrew Lang
- The Jungle Book by Rudyard Kipling
- Rilla of Ingleside by L. M. Montgomery
- The Story of the Treasure Seekers by E. Nesbit
- The Lost Prince by Frances Hodgson Burnett
- Chronicles of Avonlea by L. M. Montgomery
- Jack and Jill by Louisa May Alcott

0.4 Results

0.4.1 ANSI - Complex English - Mean 63.7 / Median 66.0

1.	MTGAP	69.24
2.	Colemak-DH	69.23
3.	Capewell	68.20
4.	Colemak	67.21
5.	Dvorak-ID	66.52
6.	Halmak	66.26
7.	Workman	65.80
8.	Asset	64.80
9.	Dvorak	63.44
10.	QGMLWY	61.07
11.	Norman	60.35
12.	QWERTY	42.80

0.4.2 ANSI - Plain English - Mean 61.7 / Median 63.7

1.	Colemak-DH	66.96
2.	MTGAP	66.63
3.	Capewell	65.79
4.	Colemak	65.15
5.	Workman	64.12
6.	Halmak	64.02
7.	Dvorak-ID	63.39
8.	Asset	63.04
9.	Dvorak	60.69
10.	QGMLWY	59.08
11.	Norman	58.68
12.	QWERTY	42.65

0.4.3 ANSI - Easy English - Mean 61.1 / Median 63.3

1.	Colemak-DH	66.09
2.	MTGAP	66.03
3.	Capewell	65.11
4.	Colemak	64.28
5.	Workman	63.85
6.	Halmak	63.68
7.	Dvorak-ID	62.82
8.	Asset	62.24
9.	Dvorak	60.30
10.	QGMLWY	58.79
11.	Norman	57.65
12.	QWERTY	42.65

0.4.4 Ergodox - Complex English - Mean 69.3 / Median 71.8

1.	MTGAP	75.10
2.	Colemak-DH	74.71
3.	Capewell	73.90
4.	Colemak	73.37
5.	Halmak	72.34
6.	Workman	71.84
7.	Dvorak-ID	71.80
8.	Asset	69.91
9.	Dvorak	68.76
10.	QGMLWY	67.25
11.	Norman	65.97
12.	QWERTY	47.04

0.4.5 Ergodox - Plain English - Mean 69.2 / Median 71.7

1.	MTGAP	74.67
2.	Colemak-DH	73.95
3.	Capewell	73.69
4.	Colemak	72.59
5.	Halmak	71.77
6.	Workman	71.70
7.	Dvorak-ID	71.59
8.	Asset	69.60
9.	Dvorak	69.00
10.	QGMLWY	66.77
11.	Norman	66.18
12.	QWERTY	48.40

0.4.6 Ergodox - Easy English - Mean 69.2 / Median 71.8

1.	MTGAP	74.45
2.	Colemak-DH	73.71
3.	Capewell	73.62
4.	Colemak	72.13
5.	Workman	71.88
6.	Dvorak-ID	71.82
7.	Halmak	71.72
8.	Dvorak	69.41
9.	Asset	69.35
10.	QGMLWY	67.03
11.	Norman	66.05
12.	QWERTY	49.12

0.4.7 ANSI - Averages - Mean 62.2 ±1.6 / Median 64.3 [1]

1.	Colemak-DH	67.4 ±1.8
2.	MTGAP	67.3 ±1.9
3.	Capewell	66.4 ±1.8
4.	Colemak	65.6 ±1.7
5.	Halmak	64.7 ±1.6
6.	Workman	64.6 ±1.2
7.	Dvorak-ID	64.2 ±2.3
8.	Asset	63.4 ±1.5
9.	Dvorak	61.5 ±1.9
10.	QGMLWY	59.7 ±1.4
11.	Norman	58.9 ±1.5
12.	QWERTY	42.7 ±0.1

0.4.8 Ergodox - Averages - Mean 69.2 ±0.1 / Median 71.8

1.	MTGAP	74.7 ±0.4
2.	Colemak-DH	74.1 ±0.6
3.	Capewell	73.7 ±0.2
4.	Colemak	72.7 ±0.7
5.	Halmak	71.9 ±0.4
6.	Workman	71.8 ±0.1
7.	Dvorak-ID	71.7 ±0.1
8.	Asset	69.6 ±0.3
9.	Dvorak	69.1 ±0.4
10.	QGMLWY	67.0 ±0.3
11.	Norman	66.1 ±0.1
12.	QWERTY	48.2 ±1.1

0.5 Discussion

According to the test method, Colemak-DH and MTGAP are the most efficient layouts while QGMLWY and Norman are the least efficient although still significantly better than QWERTY. As I see it, Colemak-DH, MTGAP, and Capewell score highly since they are based on a better understanding of ergonomics and/or an adequate analysis. The two most popular layouts, Colemak and Dvorak, are less optimized because of being based on the traditional typing technique, and in Dvorak's case, insufficient analysis due to the technical limitations at the time of its creation. However, there exist ergonomical key remappings for both layouts (Colemak-DH and Dvorak-ID) which significantly improve their scoring and require minimal relearning, making them appealing for users of those layouts. All the other tested layouts, such as Norman or Workman that claim to be better than Colemak, score lower despite being modern and having access to more ergonomics literature and more thorough machine analysis. That is because they are based on an inadequate analysis and/or a flawed understanding of ergonomics, such as the traditional typing technique that causes ulnar deviation.

[1] 95% confidence intervals from the sample standard deviations of the total averages and each layout over the 3 corpora

The total average scores of Ergodoxes vary by 0.1 for the 3 corpora while ANSI total average scores vary by 1.6. This implies that language complexity has no significant impact on typing ergonomics on Ergodox keyboards while a higher language complexity results in significantly higher scores on ANSI keyboards. This is likely due to the increased average sentence length, which results in fewer capitalised letters at the beginning of sentences. Movement of the pinkie finger to shift is penalized on ANSI keyboards while it takes minimal effort on Ergodox keyboards because of one of the thumbs resting on the shift key. The layouts score on average 7 points higher on Ergodox keyboards than on ANSI keyboards.

For more analyses and easier error-checking, raw data with the corpora, source layout files for SteveP's analyzer, and results are available at
`https://drive.google.com/drive/folders/1ao68nghEh9UoOLz_KAsGlWAN0q4Nrtbl?usp=sharing`

0.6 Conclusions

- The data strongly suggests that for the layouts tested, language complexity has no significant impact on typing ergonomics on Ergodox keyboards while on ANSI keyboards, the layouts score significantly higher in complex English.
- Even in the presence of punctuation, all of the tested alternative layouts score significantly higher than QWERTY in every corpus.
- Ergodox keyboards significantly reduce typing effort for the layouts tested and have minimal score variation.
- Out of the layouts tested, Colemak-DH and MTGAP are the most efficient alternative layouts on ANSI and Ergodox according to the typing effort model used in the analysis.

Bibliography

[1] *Dvorak Keyboard Layout*, https://en.wikipedia.org/wiki/Dvorak_keyboard_layout.

[2] *Colemak Keyboard Layout*, https://colemak.com/.

[3] *Workman Keyboard Layout*, https://workmanlayout.org/.

[4] *Ergonomic keyboards aren't healthier*, https://www.claimsjournal.com/news/national/2014/01/09/242633.htm.

[5] *Application of a genetic algorithm to the keyboard layout problem and literature review*, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6946161/.

[6] *Angle principle*, https://colemakmods.github.io/ergonomic-mods/angle.html.

[7] *Ergonomic key remappings*, https://forum.colemak.com/topic/2315-dreymars-big-bag-of-keyboard-tricks-main-topic/.

[8] *Traditional typing technique*, https://en.wikipedia.org/wiki/Touch_typing.

[9] *Negative effects of ulnar deviation*, https://pubmed.ncbi.nlm.nih.gov/20034313/.

[10] *SteveP's fork of the Patorjk analyzer*, https://stevep99.github.io/keyboard-layout-analyzer/#/about.

[11] *CarpalX analyzer*, http://mkweb.bcgsc.ca/carpalx/.

[12] *Patorjk's analyzer*, http://patorjk.com/keyboard-layout-analyzer/#/about.

[13] *CarpalX's typing effort model*, http://mkweb.bcgsc.ca/carpalx/?typing_effort.

[14] *Patorjk's analyzer source files*, https://github.com/patorjk/keyboard-layout-analyzer.