Aalto University

Dhakal, Vivek; Feit, Anna Maria; Kristensson, Per Ola; Oulasvirta, Antti

Observations on typing from 136 million keystrokes

Observations on Typing from 136 Million Keystrokes

Vivek Dhakal[1], **Anna Maria Feit**[1], **Per Ola Kristensson**[2], **Antti Oulasvirta**[1]
[1]Aalto University, Finland   [2]University of Cambridge, UK

ABSTRACT

We report on typing behaviour and performance of 168,000 volunteers in an online study. The large dataset allows detailed statistical analyses of keystroking patterns, linking them to typing performance. Besides reporting distributions and confirming some earlier findings, we report two new findings. First, letter pairs typed by different hands or fingers are more predictive of typing speed than, for example, letter repetitions. Second, rollover-typing, wherein the next key is pressed before the previous one is released, is surprisingly prevalent. Notwithstanding considerable variation in typing patterns, unsupervised clustering using normalised inter-key intervals reveals that most users can be divided into eight groups of typists that differ in performance, accuracy, hand and finger usage, and rollover. The code and dataset are released for scientific use.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords

Text entry; modern typing behavior; large-scale study

INTRODUCTION

The study presented here contributes to recent efforts to revisit the scientific understanding of typing in the era of modern computer keyboards [5, 22, 24, 32]. Historically, most studies of typing were carried out with trained typists using the touch typing system on typewriters [4, 34, 36]. However, contemporary keyboards and users are very different. Most computer users do not undertake formal training in the touch typing system, wherein each finger is responsible for pressing the keys in one or two columns. Instead, typing styles emerge and manifest themselves as highly varied strategies employing between 2 and 10 fingers [5]. Physical keyboards are still a core input device for word processing, programming, and communication, and they are used on a daily basis. Rigorous studies of patterns and predictors of typing performance have potential to improve the productivity, enjoyability, and ergonomics of computer use.

This paper presents new data and statistical analyses of modern typing. Among the empirical studies done on typing with modern keyboards, most studies of keystroke dynamics are related to biometric security (*e.g.* [6, 15, 29]). Several papers have examined typing dynamics and gesturing on small physical keyboards and mobile (soft) keyboards (*e.g.* [3, 9, 13, 25, 35, 43]). The papers considering modern typing have involved laboratory studies with small sample sizes. Nonetheless, they have provided some early insights into movement patterns. For instance, a recent study using motion tracking ($N = 30$) found that consistent use of fingers, preparatory movements, and minimal global hand motion predict high performance [5]. Even participants not using all their fingers, and not trained for the touch typing system, reached speeds comparable to, or greater than, that of touch typists [5].

In contrast to prior work with its relatively low number of participants, we report here on results for keystroke-level performance and patterns from a large-scale dataset for over 168,000 volunteers around the globe producing over 136 million keystrokes. Even if online studies of self-selected volunteers do not permit as rigorous control, large samples increase statistical power and yield better estimates and shapes of distributions [31]. They were shown to attract more varied participants than lab experiments, while yielding similar results [8, 30]. However, large-scale analyses of general typing are rare in the literature and undertaken mostly by commercial organisations, such as typing-test companies that may use results only proprietarily and not conform to standards for performance measurement.

With a goal of better understanding of modern typing behaviour and how it affects performance, we report firstly on distributions for standard metrics, including words per minute (WPM), error rate, inter-key intervals (IKIs), keypress durations, keystrokes per character (KSPC), and error corrections. We compare the speed of hands and analyse specific types of errors (omission, insertion, substitution), variations in IKIs of bigrams, and behaviours such as rollover. We then perform a detailed correlation analysis of these measures and compare groups: fast versus slow and trained versus untrained typists. This binary partitioning conceals many differences, but prior work has not provided enough ground for more detailed distinction between groups of typists. Hence, we use unsupervised clustering on keystroke-level features, to explore similarities in typing behaviour between participants. We found 8 groups of typists, which differ in hand usage, errors, and rollover behaviour, affecting their typing performance.

In summary, this paper contributes the following. First, the results expose distributions of keystroke-level metrics for a

very large global sample. Second, we present new insights into modern typing behaviour by comparing fast versus slow and trained versus untrained typists, where the large dataset gives us good estimates of the effect sizes. Third, we show which aspects of typing predict performance, reporting on new findings, such as the prevalence of rollover typing (wherein the next key is pressed down before the previous is released). Fourth, explorative analysis through clustering suggests that keystroke dynamics enable characterising typing behaviour. They can be used to cluster typists into eight groups by typing performance, accuracy, rollover, and hand usage. Finally, we discuss implications for text entry design and future research.

BACKGROUND AND RELATED WORK

Cognitive and motor aspects of typing have been a central topic of research since the introduction of the typewriter. Most scientific accounts of typing come from the 1930s–1980s, when research was carried out with mechanical or electric typewriters. Dvorak (*e.g.* [4]), Gentner (*e.g.* [7]), Shaffer (*e.g.* [36]), and Salthouse (*e.g.* [33]) used mainly professional typists deploying the touch typing strategy. Participants' weekly amount of typewriter use averaged about 11 hours [33].

These studies observed performance of around 60–75 WPM [33, 12], with average IKIs of 140 ms [36]. Letter pairs typed by fingers of different hands were 30–60 ms faster than those using fingers of the same hand [34] and about 80 ms faster than with the same finger. Error rates were found to range from 1.0% to 3.2% when typists were asked to type as quickly and accurately as possible while not correcting mistakes. Insertion errors (typing an extra letter) and omission errors (leaving one out) were more common than substitution errors (replacing one character with another) [12, 33], whereas the opposite was found for novice touch typists [12]. Others have provided more extensive reviews of empirical phenomena (*e.g.* [34, 41]). Predictive models have been proposed for typing, based on such results. Given a letter sequence, they attempt to predict the time between two keypresses. These range from theoretical (verbal) models to simulation and mathematical models. See, for example, the overview by Feit *et al.* [5].

There is no principled reason to expect that higher-level perceptual and cognitive aspects would differ dramatically between the typewriter and the modern keyboard. However, the physical properties allow for very different hand and finger movements. This could imply differences in bimanual and eye–hand coordination. Still, to our knowledge, no model exists for contemporary non-touch typists, who may use any range of fingers between 1 and 10. Datasets such as that presented here are important enablers for revisiting models.

Typing on Modern Keyboards

Only a few studies analyse movement behaviour with modern keyboards. In contrast to research on typewriters, they are conducted with less skilled typists and performance is much lower than that of professional typists reported in earlier decades [10, 20]. Logan *et al.* [24] studied how typists trade off between Fitts' law and Hick's law to find the optimal mapping from fingers to keys. They argued that trained touch typists are able to type more rapidly because they use more fingers and type more consistently, using the same finger for a given key. Rieger and Bart [32] compared touch typists and 'idiosyncratic' (non-touch) typists and found that faster typists rely less on visual information about the typing process (*e.g.* location of fingers on keyboard). This was seen for both trained and untrained typists. However, the findings were based on self-reporting by participants. Feit *et al.* [5] captured hand and finger movements for a sample of 36 users with a motion-capture system. They found that consistent use of the fingers, minimal global hand motion, and preparation of keypresses were predictors of performance, again independent of the typing strategy employed. The results suggest that typing performance is determined less by which finger is used for which key and more by other factors, such as keypress preparation and consistent finger use. Though based on a small sample, the findings encourage larger-scale data collection.

Typing on Multitouch Devices

Much of the focus on typing has shifted to mobile text entry methods, such as smartphones' and tablets' soft keyboards and the small physical keyboards used on phones. This research has aimed to improve touch accuracy, support auto-correction, or personalise keyboards [3, 9, 13, 25, 35, 43], sometimes using large-scale data collection methods [13]. However, touch keyboards are mostly operated by only 1–2 fingers. One exception is a recent study of multi-finger typing on a tabletop touch surface [37]. Average performance was found to be around 30 WPM, as compared to the 60 WPM found for typing on physical keyboards. The finger use differed from that in typing on a physical keyboard, but the authors did not further analyse this typing behaviour.

Assessment of Typing Performance

The experimental task employed in this work is transcription typing, the act of typing sequences of characters by looking at an existing written record. The process includes hand and finger movements on the keyboards, keypresses, and any key-finding strategies used [34]. The transcription process is well suited to studying movement behaviour independent of the cognitive factors related to text generation, editing, or proofreading. The text corpus is an important methodological consideration. Researchers have proposed various corpora, designed to capture statistical characteristics of the tested language, be easy to memorise, or include the full range of the alphabet [17, 26, 38]. Recently, Yi *et al.* [42] proposed word clarity as a metric for sampling test sets for touch screen keyboards. A comparison of five datasets studied by Kristensson *et al.* [22] shows differences in text entry style and performance. Our work contributes to the design of suitable test sets for typing on physical keyboards. We found that certain bigrams are more predictive of performance and are thus better indicators of typing speed.

Keystroke Patterns in User Authentication

Keystroking patterns are considered consistent for each person to a recognisable degree and hence have been proposed as a biometric security mechanism [2, 15, 18]. The dataset and findings presented in this paper can benefit research on biometric authentication.

Figure 1. The interface in the experiment. Participants were shown one sentence at a time, with progress presented at the top right.

Demographics	Result	Remark
Females	52.7%	Rest preferred not to specify
Males	41.5%	
Age: mean	24.5	75% 11–30 yrs
SD	11.2	
Countries	218	68.05% from US, 85% native language English
Took a typing course	72%	
Hours typing/day: mean	3.2	64% $<$ 2 hrs, 14% $>$ 6 hrs
SD	3.2	
Qwerty layout	98.1%	Rest local alternatives or others
Physical keyboard	43.8%	Rest on-screen (touch)
Laptop keyboard	54.15%	or small physical keyboard

Table 1. Background statistics for the participants.

DATA COLLECTION

A controlled transcription task using a modern text corpus was designed by the authors and hosted on a university server. The test was launched globally on the Internet in collaboration with a commercial organisation offering online typing courses and typing tests. In contrast to commercial web sites, our transcription task was designed in line with common practices in text entry research and analyses standard measures [40, 39]. The study represents best efforts in collection of a valid dataset on typing over the Internet, following guidelines of online study platforms [30].

Sampling and Participants

Participants in the study were self-selected from the user base of the commercial site. Our typing test was advertised on their web site as a 'scientific alternative' to their regular typing test. Users coming to the page could choose between the standard one-minute typing test with a fixed piece of text and our experiment. The user base of the company is composed mainly of younger people from the US who are interested in testing and improving their typing skills.

Table 1 summarises the demographics of 168,960 voluntary participants who completed all 15 sentences and filled in the demographic questionnaire. We observe that the volunteers in our experiment show similar demographics to the customer base the company reports: mainly teenagers and young adults, with a higher percentage of females. Many had taken part in some typing training. This sample represents people interested in testing and developing their typing skills. As is common in online volunteer studies [28, 31], ours had a high dropout rate: of ∼406,000 participants starting the study, only ∼193,000 finished the test and questionnaire. Of these, we excluded 12% as detailed below.

Task and Procedure

The task was to transcribe 15 English sentences. Participants were shown instructions stating to first read through and memorise a sentence, then type it as quickly and accurately as possible. Breaks could be taken between sentences. After acknowledging that they had read the instructions and that they gave their consent for data collection, the first sentence was displayed, as shown in Figure 1. Upon pressing Enter, the next sentence was displayed. When all sentences had been transcribed, participants had to fill in a demographic questionnaire. It asked their age and gender, country and native language, keyboard type and layout, typing experience, and number of fingers used for typing. Only then were typing results shown to the participant. The results page showed WPM and error rate in comparison to other participants in the form of a histogram with explanations of the metrics, as well as the slowest and fastest sentence and the one with the most errors. After that, participants were offered transcription of more sentences to improve the assessment. Further sentences were not included in the dataset analysed.

Our study was designed in accordance with guidelines from other online study platforms, such as LabintheWild [30]. Typing 15 sentences was considered short enough to not become tedious or exhausting, and we showed participants their progress and used simple language. Most importantly, the participants were rewarded with interesting statistics about their performance in comparison to others, which is the main motivation for participation in the study.

Materials

Sentences were drawn randomly from a set of 1,525 sentences, composed of the Enron mobile email corpus [38] and English gigaword newswire corpus [11]. These corpora are commonly used in text entry studies since they represent the language used in typing tasks with the physical keyboard. From those corpora, random sentences were chosen that contained at least 3 words and at maximum 70 characters, fewer than five numerical symbols, and only simple punctuation marks (, . ! ? '). This ensured that sentences could be typed similarly across international keyboard standards, which differ mainly in the placement of special characters (punctuation marks etc.).

Implementation

The front end was implemented as a web page using HTML, CSS, and JavaScript. The page was hosted on a university server and embedded on the company web site. Data were stored in a MySQL database, and computations at the back end were performed in Scala via the Play framework. We recorded participants' demographics, the sentences presented and transcribed, and keystroke data (timestamps for key down and up and the associated character). We collected data for a duration of three months.

Key-event Instrumentation

The timestamps were recorded via JavaScript's date.now() function. Expected precision on a regular computer is 10–15 ms. Local logging of timestamps is accurate to 1 ms with

Left hand	Right hand	Hand-alternation	Letter repetition
as, sa, er, re, sd, ds, ec, ce, ew, we, wa, aw, cr, sc, cs	lk, lo, ol, op, po, io, oi, no, on, in, ni	al, la, ak, ka, am, ma, an, na, ai, ia, so, os, sp, ps, en, ne, em, me, el, le, ep, pe	ll, cc, aa, nn, tt, ss, pp

Table 2. Categorisation of bigrams by which hand is more commonly used to type the corresponding letters.

date.now(), and modern browsers pass key-events to listeners with an overhead of 1–3 milliseconds at worst. However, OS-level threading may delay timestamping by a few milliseconds. The largest source of variability stems from the USB polling rate, which under the standard is set to 10 ms (source: usb.org HID1-11, Table E.5, page 68).

PREPROCESSING AND DATA ANALYSIS

The collected dataset was first preprocessed to remove incomplete, inaccurate, or corrupted items. As described above, we include only participants who typed all 15 sentences and completed the demographic questionnaire. Of these, we exclude those with error rate > 25% and those with technical problems or cases of participants obviously getting distracted during sentences (IKIs of >50 s). This removed ~12% of participants, a rate consistent with other online studies' [8].

Performance Measures

The computation of performance metrics follows the standard definition in text entry research [39]. If not otherwise noted, all keystrokes were included in the analysis. Computed measures were stored in the database along with the logged typing data, and further analysis used these measures:

WORDS PER MINUTE (WPM) is calculated for each typed sentence as the length of the string transcribed (in words, where one word consists of any five characters) divided by the time from the first to the last keypress (in minutes).

UNCORRECTED ERROR RATE is calculated as the Levenshtein edit distance [23] between the string presented and the transcribed string, divided by the larger number of characters between the two strings.

ERROR CORRECTIONS (%) refers to the percentage of keypresses using the Backspace (BKSP) or Delete (DEL) key during typing. Note that edit operations were not restricted to backspace; they allowed use of the mouse and arrow keys to select and delete many characters at once. From log data it is difficult to assess how often this happened.

KEYSTROKES PER CHARACTER (KSPC) is the number of keystrokes (scribed as well as non-scribed keypresses) divided by the number of characters in the final string produced.

INTER-KEY INTERVAL (IKI) is the difference in timestamps between two keypress events. For IKI-based analysis, we removed keystrokes that were typed more than 5000 ms after the previous keystroke.

KEYPRESS DURATION denotes how long a key is kept pressed, computed as the difference in timestamps between the key-down and key-up events.

Error Metrics

We used Wobbrock's TextTest software [40] to make a detailed analysis of the corrected and uncorrected errors. Given the input stream, the software computes errors in typing alphanumeric characters and the spacebar and categorises them into the following error classes:

SUBSTITUTION ERROR RATE A substitution error is one wherein a participant wrongly types one character instead of another. For example, typing the word 'road' as 'riad' substitutes the wrong letter 'i' for the letter 'o'.

OMISSION ERROR RATE An omission error (or deletion error) is one wherein a letter is completely omitted from a typed sequence. For example, typing 'commitee' for 'committee' has an omission error: a 't' is missing.

INSERTION ERROR RATE An insertion error occurs when an extra letter not present in the correct text is typed in the transcription text. For example, typing 'string' for 'sting' has an extra (inserted) 'r' and therefore is an insertion error.

The software tool was not designed to analyse large quantities of data, so we examine a stratified sample of 783 participants who employ different typing behavior. See the section on keystroke-level clustering for more details.

Differences between Hands

Our data does not include observations about which finger presses which keys. Nevertheless, we can understand the performance of individual hands by categorising the letter pairs (bigrams) as left-/right-hand or hand-alternation bigrams. The categorisation is based on the How-we-Type dataset from Feit *et al.* [5], which contains data from 50 trained and untrained typists, including information about which finger presses which key. We define a left-handm, right-hand, or hand-alternation bigram as one that for at least 90% of occurrences was typed with the left hand, right hand, or fingers of different hands). In addition, only the most frequent bigrams were included, occurring at least five times for the user. The resulting categorisation is shown in Table 2. For each user and each bigram, we first compute the *bigram IKI*, as the average over all IKI observations of that bigram typed by the user. We then use this categorisation to compute the following measures for each user: *Left IKI* = the average of bigram IKIs typed by the left hand, *Right IKI* = the average of bigram IKIs typed by the right hand, *Alternation IKI* = the average of bigram IKIs typed with fingers of different hands, *Repetition IKI* = the average of bigram IKIs for letter repetitions.

Typist Groups

For further analysis, we categorise participants into *slow* versus *fast* and *trained* versus *untrained* typists as follows: *Fast Typists* are the typists with speeds higher than that of 90% of participants (above approx. 78 WPM in our data), *Slow Typists* are those whose performance is among the slowest 10% (less than approx. 26 WPM in our data), *Trained Typists* are the ones reporting having had typing training, and *Untrained Typists* are those who reported not having taken a typing course or had training in typing.

Measure	All X̄	All σ	WPM corr r	Fast X̄	Fast σ	Slow X̄	Slow σ	Sign.	d	Trained X̄	Trained σ	Untrained X̄	Untrained σ	Sign.	d
WPM	51.56	20.20	−	89.56	9.53	20.91	4.05	●	9.38	54.35	20.80	49.00	19.73	●	0.27
IKI (ms)	238.66	111.60	-0.84	121.70	11.96	481.03	123.36	●	4.10	223.55	107.78	245.34	112.60	●	0.20
Keypr. duration	116.25	23.88	-0.29	104.49	17.38	128.99	28.85	●	1.03	118.39	23.81	115.29	23.85	●	0.13
Unc. Error (%)	1.17	1.43	-0.21	0.71	0.92	1.78	1.94	●	0.70	1.02	1.31	1.23	1.48	●	0.10
Error Correct. (%)	6.31	4.48	-0.36	3.40	2.05	9.05	6.85	●	1.12	5.90	4.40	6.50	4.50	●	0.14
KSPC	1.17	0.09	-0.40	1.10	0.05	1.24	0.14	●	1.24	1.16	0.09	1.18	0.10	●	0.15
Substitutions* (%)	1.65	1.43	-0.45	0.84	0.64	3.72	2.69	●	1.57	1.49	1.37	1.75	1.46	●	0.19
Omissions* (%)	0.80	0.57	-0.33	0.49	0.41	1.29	0.75	●	1.36	0.75	0.58	0.83	0.57	○	
Insertions* (%)	0.67	0.48	-0.15	0.50	0.33	0.80	0.46	◉	0.76	0.64	0.49	0.69	0.47	○	
Left IKI (ms)	215.23	96.80	-0.70	124.37	25.90	385.18	147.14	●	2.47	209.43	94.47	217.80	97.77	●	0.09
Right IKI (ms)	203.60	99.13	-0.68	117.24	25.03	379.81	162.40	●	2.26	195.90	97.21	207.02	99.78	●	0.11
Altern. IKI (ms)	198.26	103.95	-0.72	108.62	17.57	408.77	154.25	●	2.73	188.10	98.41	202.76	105.99	●	0.14
Repet. IKI (ms)	176.36	70.26	-0.32	144.79	27.46	230.76	126.21	●	0.94	175.64	66.20	176.68	71.98	●	0.02
Numb. fingers	6.95	2.95	0.34	8.40	2.20	5.30	3.20	●	1.16	8.00	2.46	6.50	3.00	●	0.54
Rollover ratio (%)	25.00	17.00	0.73	49.90	14.00	7.60	6.40	●	3.73	29.00	17.70	24.00	16.70	●	0.29

○ p>0.01   ● p <0.01   ● p≪0.001   X̄ : Mean value   σ : Standard deviation   d: Cohen's d value   * Results based on 783 participants

Table 3. Overview of results. At the left are the mean and SD for each measure, correlation of each measure with WPM, and indication of significance. The middle and right part compare fast with slow and trained with untrained typists, respectively. Unless otherwise denoted, statistical significance of tabulated results has been tested at the 1% level via the Mann–Whitney signed rank test.
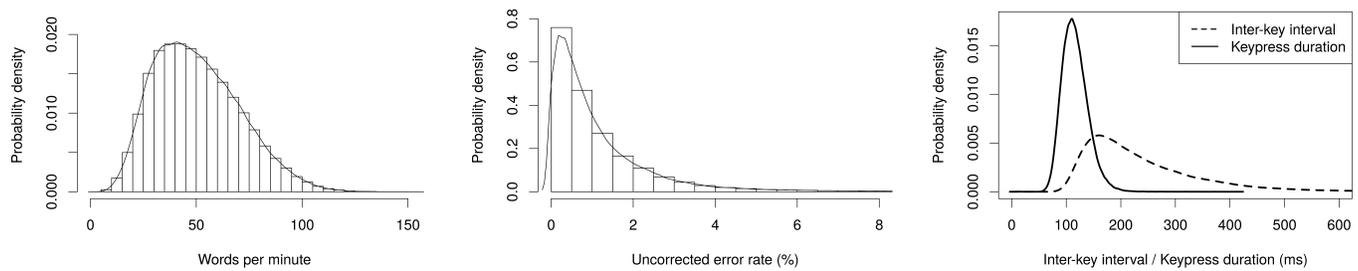

Figure 2. Histogram and density estimate of WPM, uncorrected errors, IKI, and keypress duration.

RESULTS

The final dataset includes 136,857,600 keystrokes from 168,960 participants with, on average, about 810 keypresses per participant. An overview of results is given in Table 3. It shows mean and SD for each measure and compares them for fast versus slow and trained versus untrained typists. As the data are not normally distributed and sample sizes are unequal, we use Mann–Whitney U tests to assess differences between groups. Cohen's d metric quantifies the size of the effects.

Performance Measures

Figure 2 shows the distribution of WPM, uncorrected error rate, IKI, and keypress duration over all participants.

WORDS PER MINUTE The average WPM value for participants is 51.56 ($SD = 20.2$). The fastest typists in the dataset reach speeds of even 120 WPM or more. Standard deviation is considerable even though these are self-selected participants. The distribution is not a normal-form one; it shows slight positive skewness, which is common for metrics of human performance. The distribution has a skewness of 0.513 and a kurtosis measure of -0.11. Trained typists in our sample type, on average, 5 WPM faster than untrained typists, with a similar SD, of about 20 WPM, for a relatively small difference ($d = 0.27$).

INTER-KEY INTERVALS Average inter-key interval is 238.656 ms ($SD = 111.6$). A lower bound of about 60 ms can be observed. The IKI distribution shown in Figure 2 has a skewness of 1.98 and kurtosis measure of 7.1. The differences between typists are remarkable. For fast typists, the average IKI is ∼120 ms, with a standard deviation of only 11 ms, while slow typists have an IKI of over 480 ms, sometimes as high as 900 ms, with a large standard deviation: over 120 ms. The average IKI of trained typists is only slightly less than that of untrained typists ($d = 0.2$).

KEYPRESS DURATIONS In contrast, the average keypress duration is only 116.24 ms ($SD = 23.88$) and is not shown to vary greatly even for slow typists (80–150 ms). The distribution has a skewness of 0.8 and a kurtosis of 2.36, far less than the IKI distribution has. There is a significant difference between fast and slow typists, of about 20 ms (d=1.03). Between trained and untrained typists the difference is only ∼3 ms ($d = 0.13$). Similar keypress durations have been reported in prior work (*e.g.* [21]). However, to our knowledge, the fact that keypress duration is similar across slow and fast participants has not been reported before. It is interesting, as it implies that most advances in speed are achieved elsewhere.

ERRORS AND ERROR-CORRECTION BEHAVIOUR The average uncorrected error rate of participants is 1.167% ($SD = 1.43$%).
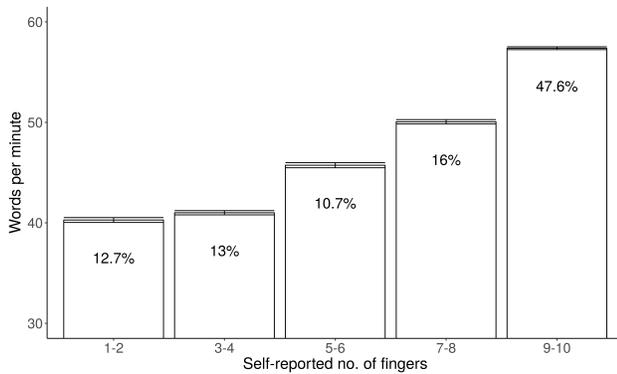
Figure 3. Self-reported number of fingers versus average WPM. The numbers denote the percentage of participants reporting using that many fingers. Error bars represent 95% confidence intervals.

The majority of participants corrected most of their mistakes, with 90% of participants leaving under 2.66% errors in the transcribed text. Slow typists leave significantly more errors uncorrected, which could indicate that they are less able to detect their mistakes. Also, trained typists leave fewer errors uncorrected ($\bar{x} = 1.02\%$) than do untrained typists ($\bar{x} = 1.23\%$). However, the difference is small ($d = 0.1$).

The average number of *error corrections* is 2.29 per sentence (6.3%, $SD = 4.5\%$), with some participants pressing Backspace or Delete up to 8.5 times per sentence, on average (99th percentile). Trained typists perform fewer error corrections than untrained typists (5.9% versus 6.5%), with large variations in both groups ($SD \approx 4.5$) This results in an average KSPC rate of 1.173 ($SD = 0.094$).

Detailed Error Measures
We observe that substitution errors (1.65%) are more frequent than omission (0.8%) or insertion (0.67%) errors. Comparing fast with slow typists, we can see that there is a particularly large difference in substitution error rate, with large variation among slow typists ($d = 1.57$), while trained and untrained typists are similar in errors. We hypothesise that substitution errors are common when participants are less consistent in typing and have a poor mental representation of the fingers' position. This result stands in contrast to prior studies in which insertion errors were commonly found to be the most frequent error type [12, 16, 33].

Differences between Hands
We found the left hand to be slightly slower than the right across all participants (average: ∼7–15 ms). Generally, bigrams typed by alternating hands were about 5–20 ms faster than those typed by only the left or right hand for fast, trained, and untrained typists. Interestingly, slow typists show the opposite effect: hand alternation is ∼20–28 ms slower. In contrast, typing the same letter twice is over 150 ms faster for slow typists, while it is 20–40 ms slower for fast typists. For both trained and untrained typists, letter repetition is faster on average and has less variation than typing other bigrams. Mann–Whitney U tests found these differences to be significant (with $p \ll 0.001$ for all comparisons).

Prior work supports the finding that the left hand is slower than the right (*e.g.* [4]), although a recent motion-capture study found the left hand to generally employ more fingers and move less on the keyboard in comparison to the right hand [5]. The hand-alternation benefit is found to be smaller than the 30–60 ms reported in many typewriter studies [34]. This is in line with prior studies of modern typists [5]. Note that the left, right, alternation, and repetition IKI values are generally smaller than the overall average IKI, since their analysis excludes many slower bigrams, such as a letter following a space (word-initiation effect [34]).

Predictors of Typing Speed
We report bivariate analyses to better understand how keystroking is related to typing speed. Pearson correlation coefficient (r) is used. Detailed results are shown in Table 3. Here, we only discuss the most important findings. In addition, we compare keystroke patterns of fast versus slow and trained versus untrained typists for frequent letter pairs (bigrams) and show that certain bigrams are more predictive of performance than others, depending on how they are typed by the hands.

Typing speed versus number of fingers
We observe a positive correlation between the number of fingers used for typing (self reported) and typing speed. Participants reported using any number of fingers from 1 and 2 to 9 and 10, with an average of 6.95 fingers ($SD = 2.95$); 47.6% of participants stated they used 9–10 fingers. The bivariate analysis in Figure 3 shows that the larger the self-reported number of fingers, the higher the typing speed ($r = 0.38$, $p < 0.001$). This relationship is non-linear. In addition, we observe fast typists to use more fingers than slow typists (averaging 8.4 versus 5.3), similarly to trained relative to untrained typists (on average, 8 versus 6.5). These results suggest that self-reported number of fingers is a predictive factor for speed and differentiates typists.

Typing speed versus errors
Not surprisingly, participants who make fewer mistakes are generally faster. We find a negative correlation between WPM and uncorrected errors ($r = -0.21$), error corrections ($r = -0.36$), and KSPC ($r = -0.4$), the last reflecting the ratio of error corrections per transcribed character. Substitution and
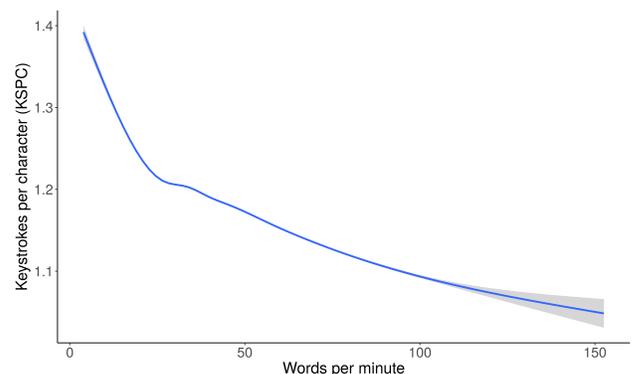

Figure 4. WPM versus keystrokes per character with 95% confidence region. Typists making and correcting fewer errors are generally faster.
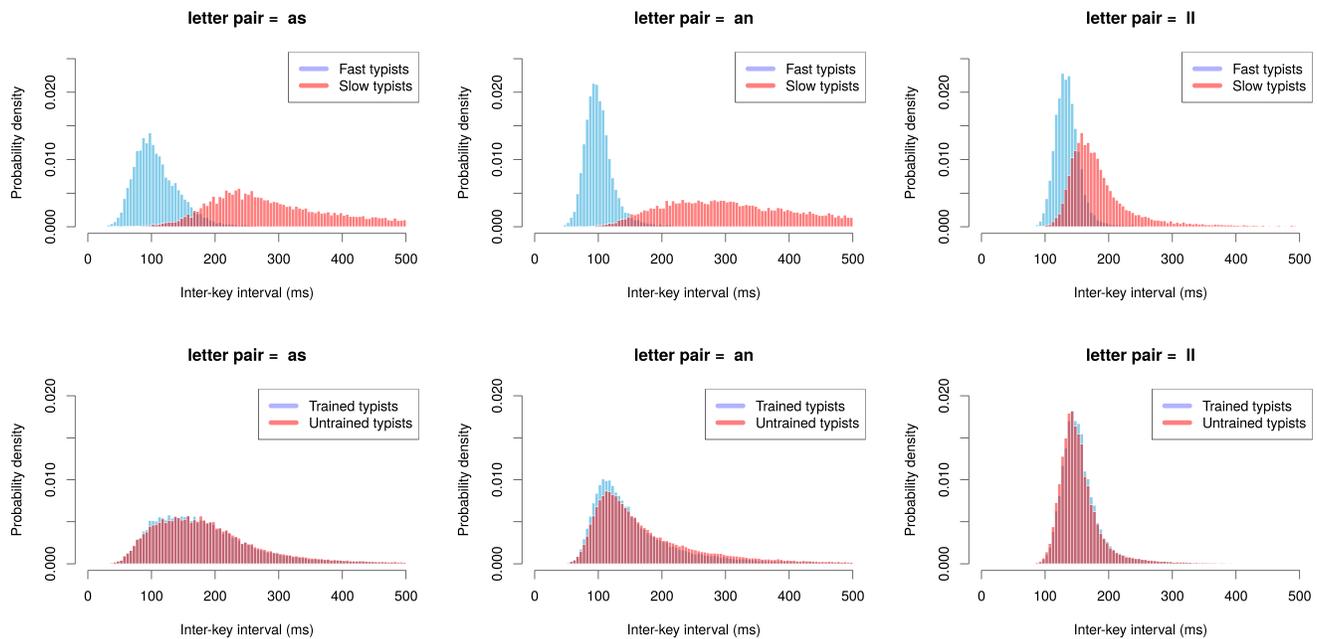
Figure 5. IKI distribution of example bigrams typed with one hand (left), hand alternation (middle), and repetition (right). Distributions differ greatly between fast and slow typists for one hand and hand alternation but are similar for repetition. Trained and untrained typists show similar distributions.

omission errors are found to correlate more with performance than do insertion errors ($r = -0.45$ and $r = -0.33$ versus $r = -0.15$), indicating that the number of insertion errors changes little with higher performance. That faster typists make fewer mistakes has been reported before (*e.g.* [12]).

The relation between WPM and KSPC is shown in more detail in Figure 4. Very slow typists (< 25 WPM) make and correct many mistakes, which results in a generally low speed (steep slope of graph). At a rate of about 25–30 WPM, the curve flattens, suggesting a group of typists who, though similar in the numbers of errors made and corrected, differ in how quickly they press keys. From about 30 WPM on, we see a linear correlation between WPM and KSPC: the faster the typists, the fewer keystrokes per character they perform; that is, they make and correct fewer mistakes. In addition to pressing additional keys, correcting errors requires locating them in the transcribed text. Therefore, as the figure shows, reducing errors by a small margin can contribute to a large gain in speed.

Typing speed versus bigram-level IKIs
We found that the IKI of letter repetitions has a lower correlation (-0.32) with WPM than that of bigrams typed by hand alternation or one hand ($\sim$-0.7), as shown in Table 3. We further explore this phenomenon by looking at the IKI distribution of specific bigrams across participants, shown in Figure 5. In the top part, we compare IKI distributions of fast and slow typists. For letter repetitions, their distributions are similarly narrow (but shifted). Variation in left-hand and alternation bigram IKIs is smaller for fast than slow typists, suggesting that they are not consistent with these bigrams.

The bottom in Figure 5 shows a similar comparison for trained versus untrained typists. We can see that for all three cate-

gories, the IKI distributions are very similar for the two groups. This indicates that variations in performance cannot be sufficiently explained by differences in which finger presses which key (*e.g.* touch typing systems versus using only 6 fingers). Instead, we must assess cognitive differences and differences in motor skill that prevent slow typists from pressing distant letter pairs as consistently as letter repetitions.

Rollover Behaviour
We report on a new phenomenon, rollover, where consecutive keypresses overlap, allowing high-performance input. While this technique is well known among keyboard manufacturers and e-sport practitioners, it has not been recognised in text entry research as a phenomenon of fast typing.

Definition of rollover ratio
Rollover refers to the technique of typing consecutive keys without releasing the earlier key. Consider, for example, the letter pair *t-h*. When using rollover, the typist first presses down *t*, then, without releasing the key, presses *h*; only then does she release *t*, and after that she releases *h*. This technique is well known to keyboard manufacturers, who have to ensure that pressing multiple keys at once is correctly sensed by the keyboard, and a keyboard supporting any number of multiple keypresses has '*n*-key rollover' [1]. However, rollover has not been studied as a typing behaviour that affects performance and was not possible on mechanical typewriters.

To quantify the percentage of keystrokes performed with rollover, we propose a new measure called **rollover ratio**. It computes the number of keystrokes typed with rollover (where the previous key is still held down at the time of the keypress) divided by the total number of keystrokes.
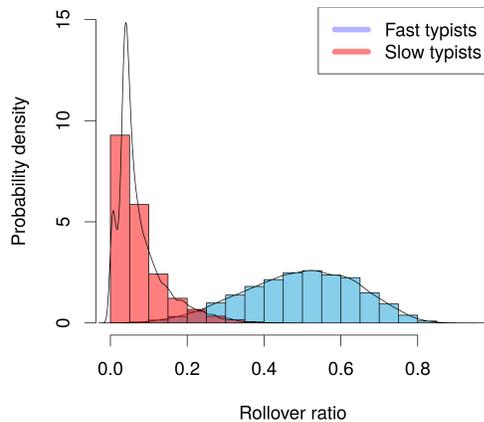
Figure 6. Distribution of rollover ratios across participants for fast and slow typists. Slow typists use nearly no rollover, whereas the majority of fast typists use rollover for 40–70% of keystrokes.

Results

The average rollover ratio is 25% ($SD = 17\%$), with most typists performing at least a small percentage of keystrokes by means of rollover. Rollover ratio is found to have a high correlation with performance ($r = 0.73$, $p < 0.001$). Figure 6 compares the distribution of rollover ratios across fast and slow typists. While slow typists use almost no rollover, the majority of fast typists use rollover for 40–70% of keypresses. Trained typists average ∼5% more rollover than untrained typists. When rollover is used, keystrokes overlap by 30 ms, on average, and up to 100 ms. The prevalence of this behaviour was surprising. To further validate it, we looked at high-speed videos of fast typists from the How-we-Type dataset, whereupon we made the same observation [5].

KEYSTROKE-LEVEL CLUSTERING OF PARTICIPANTS

Many factors predict typing performance, among them rollover, KSPC, and error rate. In biometric authentication, keystroke dynamics are used to identify users. We wanted to do the opposite: see whether one can reliably cluster typists on the basis of *similarities* in their keystroke dynamics, to find groups of typists employing similar typing behaviour.

Our goal is to summarise the major patterns of keystroking by using *unsupervised clustering*. Clustering refers to the identification of patterns whose distributions in feature space (here, normalised IKIs) are distinct. The challenge is that we have no ground truth for validating the correctness of the grouping of typists identified. We use unsupervised clustering as a tool to explore typing strategies, taking clusters' isolation as a measure for how well they can separate typing behaviour patterns of participants. Clusters are meaningful if we can interpret each group's typing behaviour.

Our clustering process revealed eight clusters. Further analysis uncovered clear differences between them with regard to performance, errors, rollover, and hand usage. This approach allows us to identify strengths and weaknesses in the participants' input strategy from keystroke data alone.

Feature Construction

Our goal was to cluster participants on the basis of their keystroke patterns, abstracting from absolute performance
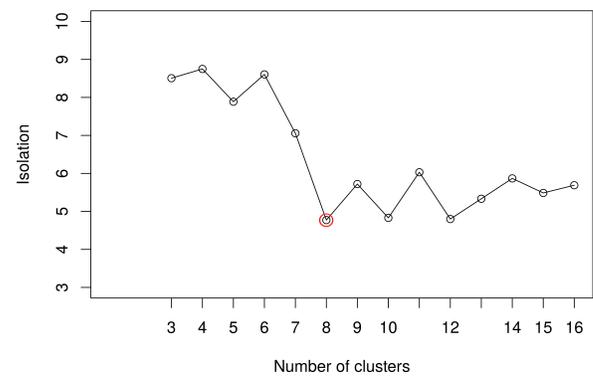

Figure 7. Isolation values for different numbers of clusters obtained by the PAM clustering method. $N = 8$ **clusters yields the best value while giving a low number of meaningful clusters.**

but capturing the typing behaviour. For each user, we constructed 38 features representing the normalised IKI for the most frequently typed bigrams, following two steps:

1. Bigram selection: Since participants typed different sentences, the dataset did not contain observations for each user and each bigram. Given the bigrams in Table 2 extended with frequent bigrams ending with a spacebar press, we chose the maximum set of bigrams and participants such that the data contained observations from at least 90% of participants. This resulted in 38 bigrams for 97.8% of participants (the rest were excluded from clustering).

2. Normalisation: For each user and each of the 38 bigrams, we constructed 38 normalised features by dividing the average IKI of each bigram by the average IKI across all 38 bigrams. This gives us a measure for how quickly a user types a bigram *relative* to the other bigrams. The normalisation abstracts from absolute performance, making it comparable across participants, and captures aspects of the typing behaviour.

WPM and other measures of participants' performance were not used as features, to focus on keystroke dynamics and to prevent explicit clusters based solely on speed groups.

k-Medoids Clustering

We chose *k*-medoids-based partitioning, specifically PAM (Partitioning around Medoids) [19], as the clustering method. Each user is represented by a data point defined by the normalised IKIs in a high dimensional space. In this space, the Euclidean distance between participants quantifies how much they differ in their keystroke dynamics in typing of different bigrams. We specifically use *unsupervised* clustering – it is unclear whether, and how, participants' typing behaviour differs, so we have no validation model and hence carry out clustering and then interpret the resulting groups. Since interpreting clusters was important, the median partitioning approach, more specifically R's clara PAM implementation [27], was used, because it produces cluster centres from among the given data points (rather than *k*-means, which converges to a set of values that may not correspond to any particular user) and is efficient enough to be used for so large a dataset.

To explore various ways of grouping participants, we ran clustering with different numbers of clusters and selected the one

Cl. #	Numb. of particip.	WPM	Rollover ratio (%)	IKI (ms)	Left IKI (ms)	Right IKI (ms)	Alternation IKI (ms)	Uncorrected (%)	Omission	Insertion	Subst.	KSPC
								Error rate (%)				
1	38,012	46.5	19.98	245.8	221.9	218.5	202.1	1.260	0.70	0.59	1.7	1.177
2	12,930	48.12	19.29	235.3	217.0	216.2	185.3	1.313	0.90	0.77	2.0	1.179
3	13,397	52.36	24.44	214.9	205.3	203.8	175.4	1.263	0.92	0.81	1.7	1.186
4	15,498	53.12	26.23	212.3	204.6	192.7	174.5	1.187	0.80	0.76	1.6	1.175
5	7,731	53.87	21.17	205.5	205.9	199.9	159.3	1.220	0.90	0.64	1.8	1.185
6	22,980	56.50	27.20	197.8	180.5	179.8	161.2	1.147	0.81	0.67	1.7	1.176
7	19,757	64.59	35.75	181.9	173.1	163.2	153.9	1.094	0.71	0.63	1.6	1.162
8	35,068	68.35	37.76	161.9	159.5	150.1	138.2	0.969	0.61	0.64	1.1	1.158

Table 4. Overview of the analysed measures differing between clusters, showing that differences in keystroke patterns affect typing performance.

that resulted in maximum isolation of clusters for further interpretation. Here, isolation refers to the measure of how compact a cluster is (that is, the average distance of any cluster member from the cluster centre) and how well-separated different clusters are (that is, the minimum distance between a cluster centre and any member of another cluster). Formally:

$$isolation_i = \frac{diameter\ of\ cluster\ i}{\arg\min_j(distance\ from\ x_i\ to\ y_j)} \qquad (1)$$

where y_j is any member of cluster $j \neq i$. The isolation value ranges from zero to infinity; the lower the better; zero means that points in the same cluster have zero distance to each other. Figure 7 compares the weighted average of the isolation values of each cluster (weighted by resulting cluster sizes) for various numbers of clusters. Eight clusters was found to result in the smallest weighted isolation value while yielding a low number of balanced and interpretable clusters.

Interpretation of Clusters
We assess the clusters by comparing differences in WPM, IKI, rollover ratio, bigram-level IKI (typed by left hand, right hand, or hand alternation), uncorrected error rate, KSPC, and errors (omission, insertion, and substitution errors). For detailed error analysis, we sampled the 783 participants who were closest to the cluster centres (within 60% of maximum distance), which best allowed interpreting differences between clusters.

An overview of the results for each cluster is given in Table 4. We can see clear differences between groups of typists for most of these measures. Note, again, that we used only keystroke dynamics (normalised IKIs) for clustering. Differences in performance are a result of differences in typing behaviour. We identified the following groups of typists:

1: SLOW, CAREFUL TYPISTS The largest and slowest group of typists (∼46 WPM). The hands are equally slow, and hand alternation is leveraged less than by others (15–20 ms faster). Only 20% of bigrams are typed with rollover. Omission and insertion errors for letters are very few in comparison to all other groups, indicating that these typists are more careful.

2: SLOW, CARELESS HAND ALTERNATORS Less than half the size of the previous group, composed of slow typists (∼48 WPM) with similar characteristics. The main difference from group 1 is that they make and correct more errors and have the highest uncorrected- and substitution-error rates. They exhibit a lower IKI by better leveraging hand alternation (∼30 ms benefit compared with either hand).

3: AVERAGE-BUT-ERROR-PRONE Average typists (∼52 WPM) who, similarly to cluster 2, make and correct many errors. They have the highest KSPC (1.186) and make the most omission and insertion errors. They compensate with a higher rollover ratio and lower IKI than groups 1 and 2 but show a hand-alternation benefit similar to group 2's.

4: AVERAGE RIGHT-HAND TYPISTS A group of average typists (∼53 WPM). The main difference from the above groups is that their right hand is much faster (∼11 ms).

5: AVERAGE HAND ALTERNATORS The smallest group of typists, with average performance (∼53 WPM). They leverage hand alternation better than even the fastest typists (40–46 ms). However, their rollover ratio is low (21%) and they make and correct many errors (KSPC=1.185), which results in only average performance.

6: AVERAGE TYPISTS Average typists with slightly higher performance (∼56 WPM), achieved through higher rollover ratio and lower IKI, with similar error rates and slightly better rollover relative to cluster 1. The hands are similarly fast, and hand-alternation benefit is comparably low (∼20 ms).

7: FAST, ERROR-PRONE TYPISTS Faster-than-average typists (∼65 WPM) with a high rollover ratio (36%). The hand-alternation benefit is comparatively low (9–19 ms), and the left hand is slightly slower than the right (∼10 ms). However, they show only slightly better error rates when compared to the groups above.

8: FAST ROLLOVERS The second-largest group, with the fastest typists (∼68 WPM) and very low IKIs (avg ∼160 ms). They show high rollover behaviour (38%) and the lowest error rates (i.e., for uncorrected and substitution errors). Similar to group 7, the left hand is slightly slower than the right.

SUMMARY
We have reported extensive statistical analysis for over 168,000 volunteer participants and more than 136 million keystrokes. The findings shed new light on typing with modern keyboards:

- The average speed is 52 WPM, but values exhibit large variance and a skewed distribution.
- While there is a large variation in IKI, keypress duration is uniformly small across typist groups: on average, 116 ms.
- Fast typists make fewer mistakes. In particular, they make fewer substitution errors, whereas insertion errors correlate less with performance.

- Rollover key-pressing is surprisingly prevalent, particularly among fast typists (40–70%), and is a strong predictor of typing speed ($r = 0.73$).
- Number of fingers (self-reported) predicts performance.
- Letter repetitions are less predictive of performance than other bigrams, indicating efficient use of hand- and finger-alternation strategy.
- Though there is considerable individual-level variation in keystroke patterns, participants can be clustered into 8 main groups, characterised by differences in typing behaviour, such as hand usage, accuracy, and error corrections.

In some respects, modern typing behaviour was found to be similar to typewriting, but we can see many differences too, particularly for hand usage. While average performance is slower, uncorrected error rates are in the same range (1–3.2%) [33, 12]. In contrast to studies of professional typists, substitution errors were more frequent than insertion and omission errors [12, 33]. We found a lower bound for IKI at 60 ms; the same was found by Hiraga *et al.* for professional typists [14]. The left hand was generally considered to be slower than the right [4]. Although present, this effect is small for modern typists (7–15 ms difference). Similarly, the hand-alternation benefit is less pronounced than reported in typewriting studies (5–20 ms versus 30–60 ms), which is in line with a prior study of modern typists [5]. On modern keyboards, using different fingers of the same hand allows for a benefit similar to that with fingers of different hands.

We find the prevalence of rollover behaviour surprising. The effect cannot be explained by our logging instrument. Rollover behaviour has a high correlation with speed ($r = 0.73$), and fast participants used rollover for 40–70% of keypresses. For fast typists it was observed irrespective of whether they were touch typists and of whether they reported using all fingers or only a few. We are not aware of previous work analysing this behaviour. Rollover behaviour needs more attention as a strategy that is not possible with traditional typewriters (it would jam the keys). Another new observation is that certain bigrams typed with different hands are more predictive than for example letter repetitions. Making use of hand alternation and finger alternation is a strategy that can be used to enter keystrokes rapidly but is not efficiently employed by slow typists. In contrast, letter repetitions are quickly typed independent of performance and show less variation in IKI.

Trained and untrained typists showed very similar typing behaviour. Slow typists make more errors, of which the majority are substitution errors. Trained and untrained typists show only small or no differences in errors. For slow typists, letter repetitions are much faster (at >150 ms) than any other bigrams, while hand alternation is not efficiently employed and can even be slower than using the same hand. We see the opposite for fast typists, for whom letter repetitions are slower (20–40 ms) and bigrams typed by hand alternation are slightly faster (10–30 ms). Slow typists use less rollover (8% versus 50%). They rely more on visual search [5, 32] and use fewer fingers, which restricts the use of rollover.

Unsupervised clustering showed that typing behaviour can be characterised via keystroke dynamics. The eight clusters differ in speed, errors, use of hand alternation, and rollover. The main difference among slow typists is in the number of errors, where careless typists demonstrate a smaller IKI and better hand alternation. Average typists differ in how they use their hands: one group is particularly fast with the right hand, and others better leverage hand alternation though their rollover ratio is lower. Fast typists gain less from hand alternation than other groups but use rollover better and make fewer errors.

Implications

The findings have high-level implications in several areas of HCI research. Modelling assumptions behind the design of text entry methods must be updated, since they have been based on a view of typing styles as rather uniform. Also, training should reconsider individuals' ways of pressing keys. Training procedures have been based mostly on the assumption of touch typing and ignored individual-specific ways of typing. Exercises could explicitly train in rollover and be personalised for typists' deficits in line with the cluster they belong to. We found that untrained typists *can* be as fast as trained typists, but trained typists in general are faster than untrained ones. It is possible that individualised training could help non-touch typists boost their performance. This implies that designers should be sensitive to the different ways people type – that is, design sensing pipelines and text entry techniques that enable rollover also on multitouch surfaces.

Limitations and Future Work

We acknowledge limitations to the generalisability of these results. The observations come from a group of self-selected participants, and 68% were from the US. Most were younger and interested in typing, with over 70% having taken a typing course, in contrast to only 43% in a prior, lab study [5]. The dropout rate is similar to other online studies' [28, 31]. Differences in sampling may explain why we found an effect for the number of fingers and speed differences between trained and untrained typists while a previous study did not [5]. We estimate our instrumentation to reach precision of about 10–15 ms per keypress. This means that all findings reported in the paper, except the left–right hand difference (7–15 ms) are likely to hold. The left–right effect should be replicated with other measurement instruments. Although edit operations (for deletion) were not limited to backspaces and deletion keys (*e.g.* mouse/keyboard multiple selection and deletion was permitted), the error analysis did not take these into account, as there was no way to track such operations. We note the need to develop instrumentation, sampling methodology, and statistical analysis methods for online typing tests.

THE 136M KEYSTROKES DATASET

The dataset (N > 168,000) and the code for the test are released at http://userinterfaces.aalto.fi/136Mkeystrokes. A subset is given in *Supplementary Material*.

REFERENCES

1. 2017. Rollover Key. (2017). `https://en.wikipedia.org/wiki/Rollover_(key)` Last accessed 2017/09/12.

2. Salil P Banerjee and Damon L Woodard. 2012. Biometric authentication and identification using keystroke dynamics: A survey. *Journal of Pattern Recognition Research* 7, 1 (2012), 116–139.

3. Edward Clarkson, Kent Lyons, James Clawson, and Thad Starner. 2007. Revisiting and validating a model of two-thumb text entry. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 163–166.

4. August Dvorak, Nellie L Merrick, William L Dealey, and Gertrude C Ford. 1936. *Typewriting behavior*. Vol. 1. New York: American Book Company.

5. Anna Maria Feit, Daryl Weir, and Antti Oulasvirta. 2016. How we type: Movement strategies and performance in everyday typing. In *Proceedings of the 2016 SIGCHI Conference on Human Factors in Computing Systems*. ACM, 4262–4273.

6. S.M. Furnell, I. Papadopoulos, and P. Dowland. 2004. A long-term trial of alternative user authentication technologies. *Information Management & Computer Security* 12, 2 (2004), 178–190. `DOI: http://dx.doi.org/10.1108/09685220410530816`

7. Donald R Gentner. 1983. Keystroke timing in transcription typing. In *Cognitive aspects of skilled typewriting*. Springer, 95–120.

8. Laura Germine, Ken Nakayama, Bradley C Duchaine, Christopher F Chabris, Garga Chatterjee, and Jeremy B Wilmer. 2012. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic bulletin & review* 19, 5 (2012), 847–857.

9. Mayank Goel, Alex Jansen, Travis Mandel, Shwetak N Patel, and Jacob O Wobbrock. 2013. ContextType: using hand posture information to improve mobile touch screen text entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2795–2798.

10. Joachim Grabowski. 2008. The internal structure of university students' keyboard skills. *Journal of writing research* 1, 1 (2008).

11. David Graff and Christopher Cieri. 2003. English gigaword LDC2003T05. (2003). `https://catalog.ldc.upenn.edu/LDC2003T05`

12. Jonathan T Grudin. 1983. Error patterns in novice and skilled transcription typing. In *Cognitive aspects of skilled typewriting*. Springer, 121–143.

13. Niels Henze, Enrico Rukzio, and Susanne Boll. 2012. Observational and experimental investigation of typing behaviour using virtual keyboards for mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2659–2668.

14. Yuzuru Hiraga, Yoshihiko Ono, and Yamada-Hisao. 1980. An analysis of the standard English keyboard. In *Proceedings of the 8th conference on Computational linguistics -*. Association for Computational Linguistics, Morristown, NJ, USA, 242. `DOI: http://dx.doi.org/10.3115/990174.990218`

15. Rick Joyce and Gopal Gupta. 1990. Identity authentication based on keystroke latencies. *Commun. ACM* 33, 2 (1990), 168–176.

16. Akiyo Kano, Janet C Read, Alan Dix, and I Scott MacKenzie. 2007. ExpECT: an expanded error categorisation method for text input. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 1*. British Computer Society, 147–156.

17. Clare-Marie Karat, Christine Halverson, Daniel Horn, and John Karat. 1999. Patterns of entry and correction in large vocabulary continuous speech recognition systems. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 568–575.

18. Marcus Karnan, Muthuramalingam Akila, and Nishara Krishnaraj. 2011. Biometric personal authentication using keystroke dynamics: A review. *Applied Soft Computing* 11, 2 (2011), 1565–1573.

19. Leonard Kaufman and Peter J. Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis.* John Wiley. `DOI: http://dx.doi.org/10.1002/9780470316801`

20. Nina Keith and K Anders Ericsson. 2007. A deliberate practice account of typing proficiency in everyday typists. *Journal of Experimental Psychology: Applied* 13, 3 (2007), 135.

21. Sunjun Kim, Jeongmin Son, Geehyuk Lee, Hwan Kim, and Woohun Lee. 2013. TapBoard: making a touch screen keyboard more touchable. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 553–562.

22. Per Ola Kristensson and Keith Vertanen. 2012. Performance Comparisons of Phrase Sets and Presentation Styles for Text Entry Evaluations. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces (IUI '12)*. ACM, New York, NY, USA, 29–32. `DOI: http://dx.doi.org/10.1145/2166966.2166972`

23. Vladimir I Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, Vol. 10. 707–710.

24. Gordon D Logan, Jana E Ulrich, and Dakota RB Lindsey. 2016. Different (Key) Strokes for Different Folks: How Standard and Nonstandard Typists Balance Fitts' Law and Hick's Law. (2016).

25. I Scott MacKenzie and R William Soukoreff. 2002. A model of two-thumb text entry. *Space* 67 (2002), 18–43.

26. I. Scott MacKenzie and R. William Soukoreff. 2003. Phrase Sets for Evaluating Text Entry Techniques. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*. ACM, New York, NY, USA, 754–755. DOI:`http://dx.doi.org/10.1145/765891.765971`

27. Martin Maechler. 2017. clara function | R Documentation. (2017). `https://www.rdocumentation.org/packages/cluster/versions/2.0.6/topics/clara` Last accessed 2017/09/18.

28. Jochen Musch and Ulf-Dietrich Reips. 2000. A Brief History of Web Experimenting. In *Psychological Experiments on the Internet*, Michael H Birnbaum (Ed.). Elsevier, 61–87. DOI:`http://dx.doi.org/10.1016/B978-012099980-4/50004-6`

29. Ngoc Tran Nguyen. 2016. Distance-based classification of keystroke dynamics. In *First International Workshop on Pattern Recognition*. International Society for Optics and Photonics, 100111E–100111E.

30. Katharina Reinecke and Krzysztof Z Gajos. 2015. LabintheWild: Conducting large-scale online experiments with uncompensated samples. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1364–1378.

31. Ulf-Dietrich Reips. 2000. The Web Experiment Method: Advantages, Disadvantages, and Solutions. In *Psychological Experiments on the Internet.*, Michael H. Birnbaum (Ed.). Academic Press, Chapter 4, 89–117.

32. Martina Rieger and Victoria KE Bart. 2016. Typing Style and the Use of Different Sources of Information during Typing: An Investigation Using Self-Reports. *Frontiers in psychology* 7 (2016).

33. Timothy A Salthouse. 1984. Effects of age and skill in typing. *Journal of Experimental Psychology: General* 113, 3 (1984), 345.

34. Timothy A Salthouse. 1986. Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological bulletin* 99, 3 (1986), 303.

35. Frode Eika Sandnes. 2005. Evaluating Mobile Text Entry Strategies with Finite State Automata. In *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices &; Services (MobileHCI '05)*. ACM, New York, NY, USA, 115–121. DOI:`http://dx.doi.org/10.1145/1085777.1085797`

36. LH Shaffer and J Hardwick. 1968. Typing performance as a function of text. *The Quarterly Journal of Experimental Psychology* 20, 4 (1968), 360–369.

37. Paul D Varcholik, Joseph J LaViola, and Charles E Hughes. 2012. Establishing a baseline for text entry for a multi-touch virtual keyboard. *International Journal of Human–Computer Studies* 70, 10 (2012), 657–672.

38. K. Vertanen and P.O. Kristensson. 2011. A versatile dataset for text entry evaluations based on genuine mobile emails. In *Proceedings of the 13th International Conference on Human–Computer Interaction with Mobile Devices and Services*. ACM, 295–298.

39. Jacob O Wobbrock. 2007. *Measures of text entry performance*. Morgan Kaufmann.

40. Jacob O Wobbrock and Brad A Myers. 2006. Analyzing the input stream for character-level errors in unconstrained text entry evaluations. *ACM Transactions on Computer-Human Interaction (TOCHI)* 13, 4 (2006), 458–489.

41. Changxu Wu and Yili Liu. 2008. Queuing network modeling of transcription typing. *ACM Transactions on Computer-Human Interaction (TOCHI)* 15, 1 (2008), 6.

42. Xin Yi, Chun Yu, Weinan Shi, Xiaojun Bi, and Yuanchun Shi. 2017. Word Clarity As a Metric in Sampling Keyboard Test Sets. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4216–4228. DOI:`http://dx.doi.org/10.1145/3025453.3025701`

43. Ying Yin, Tom Yu Ouyang, Kurt Partridge, and Shumin Zhai. 2013. Making touchscreen keyboards adaptive to keys, hand postures, and individuals: a hierarchical spatial backoff model approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2775–2784.