Predictive Models of Typing Performance in Novel Keyboards

Ioulia Simou
Computer Engineering & Informatics Department
University of Patras
Rio, Achaea, Greece
simo@ceid.upatras.gr

Andreas Komninos
Computer Engineering & Informatics Department
University of Patras
Rio, Achaea, Greece
akomninos@ceid.upatras.gr

Abstract

Novel mobile keyboard layouts require users to learn new spatial mappings, leading to performance improvements with practice. We examine how typing behaviour evolves while learning an unfamiliar virtual keyboard layout, aiming to assess the number of typing sessions required to obtain data that allow reliable predictive modelling of performance over sustained use. We find that logistic function models outperform the power function models proposed by past literature, in terms of robustness and predictive ability.

CCS Concepts

• **Human-centered computing** → **Text input**; **User studies**; **Keyboards**; **HCI theory, concepts and models**.

Keywords

Text Entry; Spatial Memory; User Studies

ACM Reference Format:
Ioulia Simou and Andreas Komninos. 2026. Predictive Models of Typing Performance in Novel Keyboards. In *Companion Proceedings of the 18th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS Companion '26), June 30–July 03, 2026, Patras, Greece.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3807968.3816000

1 Introduction and Motivation

Keyboard input remains the dominant text entry method on smartphones. The performance of mobile text entry systems is typically evaluated using quantitative performance metrics [1, 6, 8]. Despite the proposal of novel keyboard layouts, optimised to improve typing efficiency [4, 9, 10], QWERTY remains dominant across devices, due to the difficulty and high cognitive load while learning new spatial mappings. Understanding how the learning process evolves is important for estimating the theoretical performance potential of novel input techniques. Prior work has examined typing behaviour from empirical and theoretical perspectives. Empirical studies typically rely on performance metrics to describe improvements over time, while theoretical work has suggested that typing expertise develops through the gradual formation of spatial memory representations [2]. While these approaches provide valuable insights into the learning process, they do not sufficiently address the evolution of performance over time. Being able to reliably predict future performance from a minimal set of evaluation rounds is critical, since such trials are costly and fatigue-inducing for users.

We analyze the evolution of typing behaviour during repeated use of an unfamiliar keyboard layout over several days (OPTI-II [4]) and examine how different models capture and predict this evolution. Using standard text entry metrics, we characterize performance progression and evaluate how well different functional forms generalize to future performance. Based on this analysis, we provide insights into model selection and how learning behaviour can inform evaluation and the timing of design interventions.

2 Experiment Setup

Seventeen users (6 female, $\bar{x} = 23.35$ years, $\sigma = 3.03$) participated in a four-day longitudinal study. All reported extensive experience with soft keyboards, but little or no experience with non-QWERTY layouts. None were familiar with OPTI II [4]. Each day consisted of two transcription task sessions using the OPTI II layout, separated by a spatial mapping recall task of key positions. Participants performed transcription tasks on a 10.4" tablet (Samsung Galaxy Tab S6 Lite) fixed to a desk, using a custom on-screen keyboard, with key dimensions approximately 0.79×0.33 inches. The tablet setup ensured consistent interaction conditions across sessions for learning the unfamiliar keyboard layout. Sessions included four blocks of ten phrases sourced from [3, 7], and participants were instructed to copy phrases as quickly and accurately as possible using their preferred typing mode. We grouped phrases in 3 categories according to length, and in each session, we randomly sampled from each category, using short phrases for Day 1 & 2, medium length for D3 and longer phrases for D4 (phrase length $D1_{S1} : \bar{x} = 18.65, \sigma = 4.90$; $D1_{S2} : \bar{x} = 19.82, \sigma = 4.63$; $D2_{S1} : \bar{x} = 18.50, \sigma = 4.89$; $D2_{S2} : \bar{x} = 20.38, \sigma = 4.54$; $D3_{S1} : \bar{x} = 22.05, \sigma = 2.77$; $D3_{S2} : \bar{x} = 22.75, \sigma = 2.60$; $D4_{S1} : \bar{x} = 24.73, \sigma = 1.20$; $D4_{S2} : \bar{x} = 25.48, \sigma = 0.82$).

3 Results

Typing performance improved substantially across study days (Fig. 2). Mean typing speed increased from 10.83 WPM ($\sigma = 3.27$) on Day 1 to 16.10 WPM ($\sigma = 4.18$) on Day 4 (Fig. 2a). A repeated-measures ANOVA revealed a significant effect of day on typing speed, $F(3, 48) = 78.74, p < 0.001$. Bonferroni-corrected paired t-tests showed significant differences between Day 1 and all subsequent days (all $p < 0.001$), as well as between Days 2 and 4 ($p = 0.008$) and Days 3 and 4 ($p < 0.001$). No significant difference was observed between Days 2 and 3 ($p = 1.00$). Inter-key interval (IKI) also decreased across study days from 1.29 s ($\sigma = 0.40$) on Day 1 to 0.79 s ($\sigma = 0.21$) on Day 4, indicating increased typing fluency (Fig. 2b). A repeated-measures ANOVA showed a significant effect of day on IKI, $F(3, 48) = 60.15, p < 0.001$. Post hoc tests revealed significant differences for the same day pairs as observed for WPM.



This work is licensed under a Creative Commons Attribution 4.0 International License.
EICS Companion '26, Patras, Greece
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2669-9/26/06
https://doi.org/10.1145/3807968.3816000

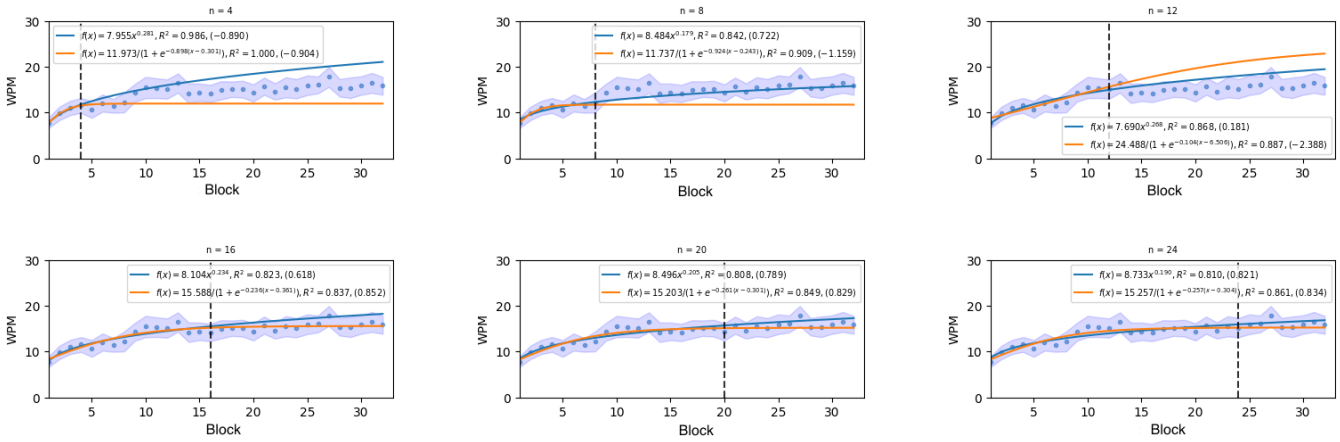
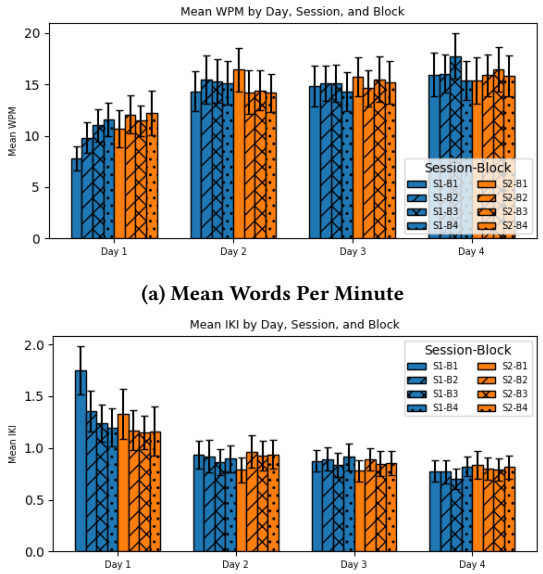


Figure 1: Fits for different sizes. Top: 4, 8, and 12 blocks. Bottom: 16, 20, and 24 blocks. In parentheses the R^2 in generalization.

Total error rates (TER) remained below 1% across all days (D1: $\bar{x} = 0.006$, $\sigma = 0.029$; D4: $\bar{x} = 0.005$, $SD = 0.024$). Normalized backspace usage remained stable (D1: $\bar{x} = 0.025, \sigma = 0.06$, D4: $\bar{x} = 0.026, \sigma = 0.05$), with no significant effect of day on backspace usage ($\chi^2(3) = 3.21, p = 0.36$), indicating that typing speed improvements were not achieved at the expense of typing accuracy.

We next examined how well different models capture and predict the observed performance increase. In prior work [4, 5] a power function model fitted across all empirical data demonstrated good R^2 values (20 and 8 sessions respectively). Predicted performance was not corroborated with further data however. More recent work suggested that the power function might be inappropriate for performance estimation [2], though results were based on simulated rather than empirical input behaviour. We examine both approaches using our dataset. Figure 1 shows model fit, across increasing training subset sizes. For small training subsets, both models generalize poorly, indicating that limited data is insufficient to reliably predict future performance. As more blocks are included, a transition becomes apparent: from 16 blocks onward, the logistic model begins to outperform the exponential function in generalization ($R^2 = 0.852$ vs 0.618). This corresponds to the point at which the slowing of performance improvements becomes evident in the typing metrics. The logistic model maintains this advantage for larger subsets, providing more accurate predictions for future performance.



(a) Mean Words Per Minute

(b) Mean inter-key interval (IKI).

Figure 2: Typing speed metrics across the four study days.

4 Discussion and Future Work

Our work examined the evolution of typing behaviour in learning unfamiliar keyboard layouts. Participants show clear improvements in typing speed while preserving accuracy, indicating genuine skill acquisition rather than a speed–accuracy trade-off. We also observe a characteristic learning pattern, with rapid initial gains followed by a slowing of performance improvements. This is reflected in speed metrics, where early sessions show substantial changes, while later sessions exhibit smaller gains. This slowing plays a key role in how well different models capture and generalize typing performance.

Power models adequately describe early learning, but fail to capture evolving performance. In contrast, logistic models provide more accurate predictions of future performance once sufficient data is available. Importantly, this transition is not observable from small subsets, suggesting that a minimum amount of evaluation trials is required for reliable modeling and prediction.

These findings have implications for both evaluation and design. Early performance data may be insufficient to predict long-term behaviour, and model selection should consider the stage of learning captured by the data. From a design perspective, the slowing of performance improvements may indicate the level of exposure after which users begin to experience slower gains. This affords designers the ability to identify opportune timing of interventions (e.g. learning aids) that may enable further improvement.

References

[1] Ahmed Sabbir Arif and Wolfgang Stuerzlinger. 2009. Analysis of text entry performance metrics. In *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)* (2009-09). 100–105. doi:10.1109/TIC-STH.2009.5444533

[2] Andreas Komninos, Ioulia Simou, and John Garofalakis. 2025. Learning to Type on Mobile Keyboards: A Theory on the Formation of Spatial Memory. 9, 5 (2025), MHCI020:1–MHCI020:27. doi:10.1145/3743720

[3] 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* (New York, NY, USA, 2003) *(CHI EA '03)*. ACM, 754–755. doi:10.1145/765891.765971

[4] I. Scott MacKenzie and Shawn X. Zhang. 1999. The design and evaluation of a high-performance soft keyboard. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (New York, NY, USA, 1999-02-01) *(CHI '99)*. Association for Computing Machinery, 25–31. doi:10.1145/302979.302983

[5] Aunnoy K Mutasim, Anil Ufuk Batmaz, Moaaz Hudhud Mughrabi, and Wolfgang Stuerzlinger. 2023. Does Repeatedly Typing the Same Phrase Provide a Good Estimate of Expert Text Entry Performance?. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2023-12-19) *(CHI EA '23)*. Association for Computing Machinery, 1–8. doi:10.1145/3544549.3585647

[6] R. William Soukoreff and I. Scott MacKenzie. 2003. Metrics for text entry research: an evaluation of MSD and KSPC, and a new unified error metric. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2003-12-05) *(CHI '03)*. Association for Computing Machinery, 113–120. doi:10.1145/642611.642632

[7] Keith Vertanen and Per Ola 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* (New York, NY, USA, 2011-05-30) *(MobileHCI '11)*. Association for Computing Machinery, 295–298. doi:10.1145/2037373.2037418

[8] Jacob Wobbrock. 2007. Measures of Text Entry Performance. In *Text Entry Systems*. 47–74. doi:10.1016/B978-012373591-1/50003-6 Journal Abbreviation: Text Entry Systems.

[9] Shumin Zhai, Hunter , Michael, , and Barton A. Smith. 2002. Performance Optimization of Virtual Keyboards. 17, 2 (2002), 229–269. doi:10.1080/07370024.2002.9667315 _eprint: https://www.tandfonline.com/doi/pdf/10.1080/07370024.2002.9667315.

[10] Shumin Zhai, Michael Hunter, and Barton A. Smith. 2000. The metropolis keyboard - an exploration of quantitative techniques for virtual keyboard design. In *Proceedings of the 13th annual ACM symposium on User interface software and technology* (New York, NY, USA, 2000-08-01) *(UIST '00)*. Association for Computing Machinery, 119–128. doi:10.1145/354401.354424