Hyper Typer: A Serious Game for Measuring Mobile Text Entry Performance in the Wild

Richard Schlögl
Christoph Wimmer
Thomas Grechenig
TU Wien, Research Group INSO
Vienna, Austria
richard.schloegl@inso.tuwien.ac.at
christoph.wimmer@inso.tuwien.ac.at
thomas.grechenig@inso.tuwien.ac.at

ABSTRACT

In this paper we introduce *Hyper Typer*, a serious Android game for collecting text entry performance data on a large scale in an unsupervised manner. Publishing the game on the Google Play Store resulted in a total of 1,917 usable transcribed phrases with 58,829 keystrokes over an eleven-week period. By analyzing the data, we demonstrate the feasibility of the method and give preliminary results regarding the overall performance and error rate of players. Moreover, the collected data allows us to compare two of the most commonly used Android soft-keyboards.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; *Empirical studies in ubiquitous and mobile computing*; *Ubiquitous and mobile computing design and evaluation methods.*

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3313035

Figure 1: Screenshot of Hyper Typer main menu

KEYWORDS

Serious Games; Games with a Purpose; Mobile Text Entry; Evaluation Methodology; Crowd Science.

INTRODUCTION

Entering text is an integral part of how we interact with our mobile devices. The use of touchscreens on modern smartphones allows for more flexibility and with it a proliferation of different text entry methods in contrast to earlier devices with physical keyboards. Both Android and iOS, the dominant smartphone operating systems in use today, allow users to download alternative keyboards and text entry mechanisms from their respective app stores to replace their default implementations. However, this abundance of choice raises the question of how to assess the utility of these text entry mechanisms and whether they provide a tangible benefit to their users.

In order to address this question and gain a better understanding of real-world text entry performance in a natural setting (i.e. on a player's personal device with their accustomed input method) we developed a game named *Hyper Typer*. Hyper Typer is a mobile typing game for Android devices with the purpose of collecting text entry performance data from players without direct supervision for exploratory data analysis.

Hyper Typer is categorized as a serious game (a game designed for a primary purpose other than pure entertainment [14]). Using serious games in research is a promising approach for turning players into voluntary participants in experiments and field studies by incentivising their participation with entertaining gameplay. For this approach to be successful, it is necessary that the tasks presented to participants can be embedded in fun and engaging game mechanics.

Related Work

Prior studies have employed serious games as an instrument in conducting HCI experiments or collecting data in the field. Henze et al. [6] used a touch-targeting game on Android to collect more than 120 million touch events from 91,731 installations. A subsequent study [7] employed a typing game to collect more than 47 million keystroke events from 72,945 installations. Rudchenko et al. [15] developed a typing game in order to provide targeting practice to users and generate training data for key-target resizing.

Serious games for research raise questions regarding the validity of results, both internal and external [3, 4], as well as ethical issues [13]. In addition, releasing and disseminating an app for research purposes through open marketplaces such as the Apple App Store or the Google Play Store is fraught with its own challenges and a number of case studies report on the varying successes and failures experienced by researchers in this endeavor [5, 12].

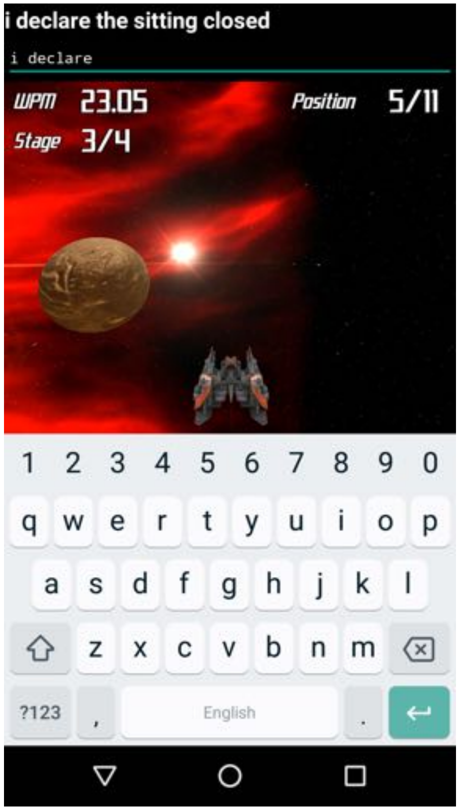

Figure 2: Screenshot of Hyper Typer gameplay

GAME DESIGN AND IMPLEMENTATION

Hyper Typer was developed in an iterative design process guided and informed by game design theory (such as the MDA framework [8] and the HCtVG model [2]), existing typing games (such as TypeRacer and ZType) and established research best practices for text entry evaluation in general and specifically on mobile devices [1, 9, 11]. While the game is suitable for conducting controlled lab experiments, its primary purpose is the ongoing collection of real-world data and performance measures of typing behaviour without supervision. To achieve this goal, the game was designed to be self-explanatory and suitable for unsupervised use and public dissemination in order to reach as large an audience as possible.

When designing serious games for research purposes, it is necessary to align the research objectives with the game design. In accordance with the HCtVG model's suggestion of deconstructing research objectives into desired outputs [2], the design of Hyper Typer and its core gameplay loop was primarily informed by the design of existing text entry experiments and associated recommendations [1, 9, 11]. The core task is the transcription of predefined phrases. All subsequent design decisions were informed by this core task and where trade-offs between experiment design and game design were necessary, decisions were favored that maintained the integrity of the research objective.

Gameplay and Setting

Hyper Typer is set in a science fiction setting where players are challenged to compete in a spaceship race. This setting was chosen because racing games are generally a mostly non-violent genre, yet create a sense of urgency and allow for competition among players. While Hyper Typer is a single player game, online highscore leaderboards encourage a sense of competition and mastery among the player base. In addition, the game features achievements (via Google Play Services) and unlockable ship models as an additional motivational affordance.

The game is structured around competitive races against computer controlled opponents, where each race consists of four individual rounds. A single round corresponds to the transcription of a single phrase. Players are presented with a phrase on the top of the screen and are challenged to transcribe the presented text as quickly and error-free as possible (the core mechanic of the game) in order to beat their opponents and achieve a highscore (see Figure 2).

Hyper Typer presents players with their familiar Android device keyboard as configured in the Android system settings. This is in contrast to other studies [7], which provide their own custom keyboard implementation for text entry. While the development of a custom keyboard provides study administrators with more fine grained control over the text entry methods used by the player, it would be contrary to our stated purpose of collecting data about realistic typing behaviour. Moreover, measures can be gathered for a variety of text entry mechanisms available on the player's device.

Table 1: Results of all valid transcribed phrases ($n = 1,917$)

	WPM	CER	UER	TER
Mean	33.51	5.154	3.587	8.741
SD	12.18	8.968	8.951	12.26
	KSPC	**PC**	**UB**	**WB**
Mean	1.842	0.5042	0.8775	0.1225
SD	3.577	0.4363	0.1598	0.1598

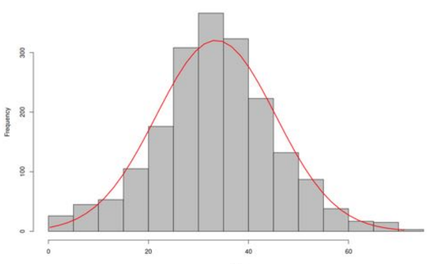

Figure 3: WPM Distribution

Study Setup and Remote Configuration Options

Hyper Typer consists of a Unity container embedded in an Android application. In order to speed up development, 3rd party assets from the Unity Asset Store were used. Experiment parameters are remotely configurable through Firebase without requiring the player to update the app. The remote configuration can be identical for all participants or partitioned as an A/B test.

The game contains phrase sets in four languages (English, German, French and Spanish) sampled from transcriptions of the European Parliament using the method described by Leiva and Sanchis-Trilles [10]. The language of the presented phrases is matched to the device language setting, with English chosen as a fallback if no matching phrase set is available. Additional phrase sets can be added remotely and are downloaded in the background before a new game is started. Experiment results are sent via REST API and stored in a PostgreSQL database. All data is transferred in the background without player interference. In case of the device being offline, data transmissions are queued in the background and resumed at a later time when connectivity has been restored.

Measures and Data Collection

Before displaying the main menu for the first time, the game informs players about the research purpose and collected data in order to ensure GDPR-compliance and gain informed consent from players. The game collects data about device characteristics, typing performance and player behaviour. A unique installation identifier is created when first starting the app. In addition, the device model, a keyboard identifier, Android SDK version and whether the game was paused are logged for each game played. No personal information is collected by the game.

The game calculates the following performance measures for text entry speed, error rate metrics and efficiency (for a detailed description see [11]): Words per Minute (WPM), Minimum String Distance (MSD), Keystrokes per Character (KSPC), Total Error Rate (TER), Corrected Error Rate (CER), Uncorrected Error Rate (UER), Correction Efficiency (CE), Participant Conscientiousness (PC), Utilised Bandwidth (UB) and Wasted Bandwidth (WB). In addition, the game stores a time-stamped log of all keyboard input events for detailed character-level analysis.

PILOT STUDY AND PRELIMINARY RESULTS

Hyper Typer was initially tested in an Amazon Mechanical Turk experiment. 10 participants were asked to download and play the game five times with each game consisting of four rounds. No explanation of the game mechanics was given in order to gather insights on the comprehensibility of the game design. All participants were able to complete the task within 18 minutes or less. 5 participants provided feedback which was uniformly positive and indicated no problems of understanding or technical issues.

Table 2: Performance per Keyboard

	Android Default	Samsung	
n	1,062	188	
	WPM		p
Mean	35.82	31.33	<.001
SD	10.81	11.89	
	TER		p
Mean	6.444	10.90	<.001
SD	9.337	15.12	

Figure 4: QRCode link to Hyper Typer on the Google Play Store [16]

Following the successful pilot test, Hyper Typer was released publicly on July 13th, 2018. Between its release and October 1st, 2018, 893 games were played on 275 distinct installations. Of those 275 installations, only 47 reached a final score higher than 0 points in 504 games. 6 games were interrupted and later resumed and therefore excluded from the results. To remove garbage input, transcribed phrases with an individual score of 0 were also filtered out. This results in a total of 1,917 usable transcribed phrases with 58,829 keyboard input events (Table 1).

Each participant played 10.7 games (median=4, SD=24.04). The high standard deviation is the result of one participant playing 144 games in total. The WPM per transcribed phrase is normally distributed with a mean of 33.51 (SD=12.18) (Figure 3).

A comparison (Mann-Whitney U Test) between the two most-used keyboards among players of the game showed that players using the *Android Default Keyboard* had a significantly higher text entry speed WPM and lower total error rate TER compared to the *Samsung Keypad* (Table 2).

DISCUSSION AND FUTURE WORK

Preliminary results demonstrate the feasibility of measuring text entry performance with a serious game on a large scale. Certain aspects of the study setting, such as device and keyboard used for text entry, hand position (e.g. one or two handed, left or right handed), participant conscientiousness or context of use are outside the control of study administrators. However, in order to ensure the quality and integrity of the collected data, certain collected usage characteristics such as device identifier, keyboard type or pauses in gameplay can be used to reject invalid or inapplicable results on a case by case basis.

While the game is finished and has so far proven to be fit for its intended purpose, further dissemination, data collection and exploratory data analysis remain ongoing efforts. In addition, the configuration options provided by the app allow its use for exploring more focused research questions, such as the effects of auto-correction and auto-completion or the influence of screen size on performance.

The amount of garbage data entered warrants further investigation. This could be a result of bad usability or disinterest of participants. In addition, specific game design choices such as incentive structures and scoring mechanisms could distort the typing behaviour of players. To better understand the effect of these issues we plan a comparative study of data collected from unsupervised players and from a controlled lab experiment in the future. Moreover, we hope to publicly release the raw data in the spirit of open research in the future.

REFERENCES

[1] Steven J. Castellucci and I. Scott Mackenzie. 2013. Gathering Text Entry Metrics on Android Devices. In *Proceedings of the International Conference on Multimedia and Human-Computer Interaction - MHCI 2013*. 1–7.

[2] Corey Clark, Ira Greenberg, and Myque Ouellette. 2018. A model for integrating human computing into commercial video games. *2018 IEEE 6th International Conference on Serious Games and Applications for Health, SeGAH 2018* (2018), 1–8. https://doi.org/10.1109/SeGAH.2018.8401316

[3] Leah Findlater, Joan Zhang, Jon E. Froehlich, and Karyn Moffatt. 2017. Differences in Crowdsourced vs. Lab-based Mobile and Desktop Input Performance Data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. 6813–6824. https://doi.org/10.1145/3025453.3025820

[4] Niels Henze and Martin Pielot. 2013. App stores: external validity for mobile HCI. *Interactions* 20, 2 (2013), 33–38. https://doi.org/10.1145/2427076.2427084

[5] Niels Henze, Martin Pielot, Benjamin Poppinga, Torben Schinke, Susanne Boll, and Niels Henze. 2011. My app is an experiment: Experience from user studies in mobile app stores. *International Journal of Mobile Human Computer Interaction (IJMHCI)* 3, 4 (2011), 71–91.

[6] Niels Henze, Enrico Rukzio, and Susanne Boll. 2011. 100,000,000 Taps: Analysis and Improvement of Touch Performance in the Large. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. 133–142.

[7] 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 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. 2659–2668. https://doi.org/10.1145/2207676.2208658

[8] Robin Hunicke, Marc LeBlanc, and Robert Zubek. 2004. MDA: A Formal Approach to Game Design and Game Research. In *Workshop on Challenges in Game AI*. 1–4. https://doi.org/10.1.1.79.4561

[9] 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*. 29–32. https://doi.org/10.1145/2166966.2166972

[10] Luis A. Leiva and Germán Sanchis-Trilles. 2014. Representatively memorable: sampling the right phrase set to get the text entry experiment right. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. ACM Press, Toronto, Ontario, Canada, 1709–1712. https://doi.org/10.1145/2556288.2557024

[11] I. Scott MacKenzie and Kumiko Tanaka-Ishii. 2010. *Text entry systems: Mobility, accessibility, universality*. Elsevier.

[12] Donald McMillan, Alistair Morrison, Owain Brown, Malcolm Hall, and Matthew Chalmers. 2010. Further into the wild: Running worldwide trials of mobile systems. In *International Conference on Pervasive Computing*, Vol. 6030 LNCS. 210–227. https://doi.org/10.1007/978-3-642-12654-3_13

[13] Donald McMillan, Alistair Morrison, and Matthew Chalmers. 2013. Categorised ethical guidelines for large scale mobile HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. 1853. https://doi.org/10.1145/2470654.2466245

[14] David R. Michael and Sandra L. Chen. 2005. *Serious Games: Games That Educate, Train, and Inform*. Muska & Lipman/Premier-Trade.

[15] Dmitry Rudchenko, Tim Paek, and Eric Badger. 2011. Text text revolution: A game that improves text entry on mobile touchscreen keyboards. In *International Conference on Pervasive Computing*., Vol. 6696 LNCS. 206–213. https://doi.org/10.1007/978-3-642-21726-5_13

[16] Richard Schlögl, Christoph Wimmer, and Thomas Grechenig. 2018. Hyper Typer. Google Play Store. Retrieved February 14, 2019 from https://play.google.com/store/apps/details?id=at.hypertyper.