In D. Reitter & F. E. Ritter (Eds.), Proceedings of the 14th International Conference on Cognitive Modeling (ICCM 2016). University Park, PA: Penn State.

Exploring the Effects of Different Text Stimuli on Typing Behavior

Ignacio X. Domínguez (ignacioxd@ncsu.edu), Jayant Dhawan (jdhawan2@ncsu.edu), Robert St. Amant (stamant@csc.ncsu.edu), David L. Roberts (robertsd@csc.ncsu.edu)
Department of Computer Science
North Carolina State University
Raleigh, NC 27695-8206 USA

Abstract

In this work we explore how different cognitive processes affected typing patterns through a computer game we call *The Typing Game*. By manipulating the players' familiarity with the words in our game through their similarity to dictionary words, and by allowing some players to replay rounds, we found that typing speed improves with familiarity with words, and also with practice, but that these are independent of the number of mistakes that are made when typing. We also found that users who had the opportunity to replay rounds exhibited different typing patterns even before replaying the rounds.

Keywords: Typing; Cognitive Model; Keyboard; Input Device; Cognition; Speed-Accuracy Trade-off; Speed; Accuracy; Inter-keystroke Interval; Typing Mistakes; Human-Computer Interaction; Human Information Processing; Human Behavior.

Introduction

In this paper we present early exploratory work toward creating cognitive models of interaction from patterns in typing that can be applied to, among other things, identifying the cognitive processes at play when users are typing on keyboards. In particular, we focus on the use of a keyboard as a window into interaction patterns that are reflective of the user's cognitive state. By mining for patterns on the usage of input devices we aim to unobtrusively obtain a snapshot of users' perception and decision-making processes in real-time.

To explore typing patterns and their relationship to cognition we created a computer game that involved typing words of different lengths with varying *word shapes* (Bouwhuis & Bouma, 1979). Some participants were able to replay rounds. We recorded typing speed and accuracy expecting an improvement as the rounds were replayed, as well as better speed and accuracy while typing words more similar to dictionary words. By changing the nature of the words being typed, we were able to alter the cognitive process required to type them, allowing us to measure how the differences in cognition are reflected in typing patterns. Our overarching goal beyond this paper is to be able to create models of interaction that would allow real-time detection of the user's cognitive state across a wide variety of tasks and interfaces.

Related and Prior Work

Computational Cognitive Modeling

When typing or pointing at targets in a graphical user interface, users exhibit distinctive patterns in the timing of their keystrokes (Monrose & Rubin, 1997) and the movement of the mouse (Ahmed & Traore, 2007). At a more abstract level, human decision-making has also been studied. A range of "microstrategies" (Gray & Boehm-Davis, 2000) applied to low-level HID usage have been identified. Microstrategies are characteristic choices that users make, without extensive deliberation, between different actions to achieve their goals.

A representative example of research on microstrategies is due to Gray and Fu (2004), where participants were given a task to perform in a user interface that contained an information box, with a variable cost of accessing that information in different conditions: the information could be permanently visible or it could require a mouse click (with a temporary lockout) to see. They found patterns in completion times, error rates, and decisions made by participants across the conditions, which could be explained in terms of trade-offs between perceptual/motor and memory retrieval effort. Participants' behaviors depended in subtle ways on cognitive biases (*e.g.*, a preference for "knowledge in the head" rather than perfect knowledge in the world).

Models of Typing

Transcription typing has been well-studied, with some work looking at how typing speed varies with unfamiliar material. Salthouse (1986) observes 12 "basic phenomena" about typing, one of which describes the reduction in the typing speed when the typist is presented with random sequences of letters.

John (1996) introduced the TYPIST model that "can be used to make quantitative predictions of performance on typing tasks". This model is based on the Model Human Processor (MHP) (Card, Moran, & Newell, 1986). TYPIST applies the MHP to human typing tasks for skilled transcription typists in order to quantitatively predict the performance of the typists. It processes text at the level of *chunks*, which could be words, syllables, or letters. TYPIST is applied to several common typing tasks, and its predictions of the performance of the typists come to within 20% of empirical measurements.

While previous work has focused on skilled or "expert" typists, little work has explored typing patterns of average users. While Feit, Weir, and Oulasvirta (2016) recently explored the mechanics and strategies of everyday typists, the cognitive processes involved in typing different types of content remains largely unexplored.

Method

Because of the exploratory nature of this work, we focused primarily on establishing internal validity. Our main goal was to get insight into how the cognitive processes associated with typing change in relation to changes in what is being typed and previous exposure to the material. Our approach focuses on the use of computer games that elicit examples of different typing behaviors. In particular, we designed and implemented *The TypingGame*—a casual game that we present below.

Using a computer game provides some advantages at this early stage of our research. First, because game mechanics often result in changes to the details of tasks, users tend to be more accepting of changes to an interface or expectations on their performance. Second, and perhaps more importantly, computer games provide motivational context. In order to get reasonable data, users had to have an incentive to perform the task well. The "gameification" of the task enabled us to study users under experimental conditions with relatively higher engagement when compared to a stand-alone typing task.

Target Population and Sampling

We targeted computer users of at least 18 years of age, and recruited using a combination of convenience and snowball sampling. We advertised our study primarily to the Computer Science student body at the authors' institution, but also posted fliers on nearby bulletin boards. Participants were offered a base compensation of \$5.00 and a maximum of an additional \$2.00 for each game round they completed, for a total maximum of up to \$25.00 based on their gameplay performance. Interested individuals were asked to sign up online for an available time slot and location.

Our sample consisted of 43 participants, of which 14 were female and 29 were male. Before the game, we asked the participants to rate their typing skills by choosing one of these options: *Beginner*, *Intermediate*, *Advanced*, or *Expert*. Of the females, 8 reported their skills as intermediate, and 6 as advanced. In the case of the male participants, 2 reported their skills as beginner, 14 as intermediate, and 13 as advanced. The average age of the female participants was 23.57 ($SD = 2.53$) years, and for the males, it was 23.90 ($SD = 2.13$) years.

The Typing Game

Our implementation of *The Typing Game* was written in Adobe Flash CS5.5 and was designed to run on a Web browser. The goal of the game is to type words that are shown on a 4×4 game board grid as fast as possible. Sets of between 1 and 4 words, initially shown on the first row of the grid, one per column, drop down one row at periodic intervals until they are correctly typed or fall off the board. Words that are correctly typed immediately disappear from the board. If a mistake is made while typing a word, the word resets and must be typed again starting with the first letter.

The Game Experience Upon launch, our game randomly assigned the player to one of three experimental conditions. To ensure that the game screen had input focus and that the keyboard input was received by our game interface, the first screen prompted the player to press the SHIFT-N key combination to begin and presented a small demographics survey. The following screen presented a small survey that asked about the player's background and typing habits. Next, the game asked the participant to type the sentence "*the quick brown fox jumps over the lazy dog*". This sentence was used to ensure that the keyboard was working properly. Once this sentence was correctly typed, the player was prompted to press the SHIFT-N key combination to proceed to an in-game tutorial. The player was asked to press the space key to begin the tutorial, which started by explaining the game mechanics in an interactive manner prompting the player to type the word "go" in order to move to the next screen. This illustrated how words were removed from the game board once they were correctly typed. The next screen in the tutorial illustrated how sets of words would drop from the row on every time interval, and how drops affected scoring.

Table 1: Description of the rounds in our Typing Game.

Round	Round	Word Length	Word Type
Practice	Practice	Short	Dictionary word
1	DictM	Medium	
2	ShapeS	Short	Transposed letters preserving word shape
3	ShapeM	Medium	
4	ShapeL	Long	
5	NoShapeS	Short	Transposed letters breaking word shape
6	NoShapeM	Medium	
7	NoShapeL	Long	
8	RandS	Short	Random letters
9	RandM	Medium	
10	RandL	Long	

Next, the game introduced participants to a practice round that accurately simulated the mechanics that the player would experience in the game rounds. In order to advance to the game rounds, the player was required to earn at least \$1.70 during the practice, and was required to replay the round until she did. The money earned during the practice round did not count towards the participant's final compensation.

To ensure that players were ready, each round had a staging screen that prompted the participant to press the space key to begin. After each round, a summary screen presented the round number, the amount earned, and a prompt to press the SHIFT-N key combination in order to proceed (some conditions also displayed a prompt to press the SHIFT-R key combination to replay the round, as described below). In addition to the practice round, participants completed a total of 10 game rounds, which did not require a minimum score.

Rounds A single game round contains multiple word sets that initially appear on the first row, but on different columns, of the game board grid. Our game consisted of 10 rounds varying the type of words and their length (see Table 1).

We designed our rounds with four types of words, all in lowercase: 1) dictionary words (*e.g.*, "quit") , 2) dictionary words with one or more transposed letters, preserving the general shape of the word (*e.g.*, "tiem" for time), 3) dictionary words with one or more transposed letters, breaking the general shape of the word (*e.g.*, "gluf" for gulf), and 4) words composed of random letters, filtering out common bi-grams and tri-grams to avoid confounding our variables. The idea behind the differences in word choices was to explore how the similarity of the word being typed to a real word affected the typing patterns. For the same reason, our rounds had different word lengths (short, medium, and large, as shown in Table 1, with 3-4, 4-5, and 5-6 characters, respectively).

Scoring Every round begins at the highest score (\$2.00) and decreases by \$0.05 (until it reaches \$0) for every time a set of untyped words drops down one row. For a player to earn the maximum score, she has to type every word correctly while they are still on the first row. The amount to be earned for a round is displayed at the bottom right of the game screen and is updated as the words drop.

Depending on the experimental condition, some players had the option to replay rounds by pressing the SHIFT-R key combination during a round's summary screen. The score earned for a round would be the one obtained on the last replay of that round, regardless of whether it was lower or higher than the score obtained in previous attempts.

Visual Design The 4×4 game board grid has a black background, where each cell is 200px wide and 100px tall. A cell with an untyped word will have a gray background. Every word uses the Consolas font in 18 point. The color of the font is initially black, but as a word is typed, the color of correctly-typed letters changes to a dark gray to show progress.

Experimental Procedure

The researchers asked participants to meet them at a designated room during a time slot previously agreed upon. After providing informed consent, participants were given the opportunity to ask questions before moving on to the data collection phase. At this time, the researchers would instruct participants to sit in front of a computer that was previously set up to run our game using the Google Chrome browser in full-screen mode. This computer was instrumented with a USB Microsoft Wired Keyboard 600 configured with US American visual and functional keyboard layouts. The researchers asked participants to notify them once they reached the final screen of the game and stepped out of the room, leaving the participants alone with no distractions.

Once they completed the game, participants would notify the researchers who would then record the participant's earnings and a unique game-generated code from the last game screen onto a paper form that participants would later use to collect their compensation. The purpose of this last step was to avoid associating participants with the data that was collected from their participation.

Experimental Conditions

Participants were randomly assigned to one of three experimental treatments.

- **Replay not allowed:** Participants were <u>not</u> allowed to voluntarily replay any rounds. The practice round could only be replayed until the minimum score of \$1.70 was obtained. The summary screen of every round only allowed participants in this treatment to advance to the next round.
- **Replay encouraged:** Participants were allowed to voluntarily replay any round an unlimited amount of times, including the practice round after the minimum score of \$1.70 was obtained. The summary screen of every round showed both the key combination to press in order to advance to the next round and the key combination to press in order to replay the round.
- **Replay allowed:** Participants were allowed to voluntarily replay any round an unlimited amount of times, including the practice round after the minimum score of \$1.70 was obtained. The summary screen of every round showed both the key combination to press in order to advance to the next round and the key combination to press in order to replay the round, but the latter was displayed as if it were inactive (grayed out) despite being functionally equivalent to the **replay encouraged** treatment.

We found that participants in the *Replay allowed* treatment never attempted to replay a game round and therefore behaved in the same way as participants in the *Replay not allowed* treatment. We believe that displaying the replay prompt as inactive was enough to make participants believe that they did not have the ability to use that feature. For the purpose of our analyses, we will treat all participants in these two treatments as a single group. We will refer to these groups as the **replay** (16 participants) and **no replay** (27 participants) conditions based on whether they voluntarily replayed rounds or not, respectively.

Logs and Analytics

We had three independent variables in our experiment: 1) the *Round* of our game being played, which modified the length and type of words that players had to type, 2) the *Condition*, which dictated player's ability to replay rounds, and 3) the *Attempt*, which indicates how many times the round is being replayed. In the case of the no replay condition, the value of the *Attempt* of a game round is always 1.

Our implementation of *The Typing Game* captured the "key down" and "key up" keyboard events, causing each keystroke to be recorded as two events. In addition to the key that generated the event, our game also collected a timestamp, with millisecond precision, of when each event occurred. Each key event was also associated to the round or screen active when it occurred, to any word on the board to which it may have corresponded, whether the keystroke was correct or not, and whether it completed a word on the board. These low-level data allow us to calculate higher level metrics. In particular, in this paper we define the following analytics:

- **Inter-keystroke interval (IKI):** the number of milliseconds elapsed between the "key down" events of each contiguous pair of keystrokes in a correctly-typed word. For the purposes of this metric, we excluded events from words that were typed with mistakes.
- **Number of mistakes:** the count of keystrokes during a round that did not clear the game board, or that did not result in the board being one character closer to being cleared. For the purposes of this metric, we did not count whitespace characters as mistakes.

The IKI is a common metric for typing speed (Salthouse, 1986), while the number of mistakes is a natural metric for typing accuracy. We will refer to *typing speed* as the inverse of the IKI, where a smaller IKI represents an increase in speed (and vice versa), and to *typing accuracy* as the inverse of the number of mistakes made, where fewer mistakes indicate a higher accuracy (and vice versa).

Formally, our hypotheses are:

H1: Practice increases speed – The average IKI in a round will be smaller when replaying.

H2: Practice increases accuracy – The average number of mistakes in a round will be smaller when replaying.

H3: Familiar words are typed faster – The average IKI of a word will be smaller the closer the word is to a dictionary word.

H4: Familiar words are typed more accurately – The average number of mistakes made when typing a word will be smaller the closer the word is to a dictionary word.

Analysis and Results

To ground the internal validity of our study with respect to both speed and accuracy, we compared the first attempt of the practice round between both the replay and no replay conditions. Because the game experience for both conditions is identical at this point in the game, we expected no substantial difference between the two. To evaluate the significance of the difference in the average IKI between the replay ($M = 164.32, SD = 92.45$) and no replay ($M = 163.62, SD = 121.77$) conditions on the first attempt of the practice round, we conducted a Welch's independent-samples t-test, which revealed no significant difference in speed ($t(1170.2) = -0.12995, p = 0.8966$). To evaluate the significance of the difference in the average number of mistakes between the replay ($M = 5.44, SD = 5.51$) and no replay ($M = 5.78, SD = 5.89$) conditions on the first attempt of the practice round, we conducted a Welch's independent-samples t-test, which revealed no significant difference in accuracy ($t(33.331) = 0.19074, p = 0.8499$).

Having established the first attempt of the practice round as a valid baseline across our experimental conditions, we used the individual player's averages of IKI and number of mistakes on this attempt of this round to normalize their game rounds' IKI and number of mistakes, respectively, by dividing the measured value by the average. We use these normalized values for the rest of our analyses. Descriptive statistics for IKI and number of mistakes made in each round by attempt are shown in Table 2 and Table 3, respectively.

Improvement with Practice

This part of the analysis focuses on the replay condition as it was the only one that allowed replaying rounds. Even though participants in the replay condition were allowed to replay as many times as they wanted, the most participants replayed a single round was 8 times. However, because at most 3 participants replayed a single round more than 4 times, we decided to focus on the first 4 attempts in our analysis.

Our hypothesis **H1** expects there to be an improvement in speed as rounds are replayed. We conducted a factorial ANOVA to examine the effects of Attempt and Round on the IKI. The results yielded a main effect for the attempt ($F(1, 15471) = 102.4765, p < 0.001$), indicating that the typing speed of participants significantly increased (*i.e.,* the IKI decreased) the more rounds were replayed. The main effect of the round was also significant ($F(9, 15471) = 102.4765, p < 0.001$). The interaction effect was non-significant ($F(9, 15471) = 0.8436, p > 0.1$). This results is consistent with our hypothesis **H1**.

Table 2: Normalized mean and standard deviation of the inter-keystroke interval of participants on the "replay" condition on each of the first four attempts of every round.

	Attempt 1		Attempt 2		Attempt 3		Attempt 4	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
DictM	1.03	0.58	1.02	0.61	0.96	0.63	0.96	0.41
ShapeS	1.22	0.69	1.18	0.60	1.12	0.77	1.15	0.48
ShapeM	1.29	0.90	1.24	0.95	1.13	0.70	1.05	0.55
ShapeL	1.52	1.07	1.40	0.92	1.36	0.86	1.25	0.75
NoShapeS	1.21	0.65	1.15	0.55	1.06	0.50	1.05	0.60
NoShapeM	1.37	0.94	1.26	0.72	1.19	0.72	1.11	0.68
NoShapeL	1.46	0.99	1.42	0.82	1.21	0.64	1.22	0.75
RandS	1.41	0.89	1.25	0.72	1.18	0.55	1.24	0.80
RandM	1.68	1.36	1.58	1.03	1.37	0.85	1.54	1.50
RandL	1.95	1.49	1.91	1.45	1.77	1.14	1.56	1.03

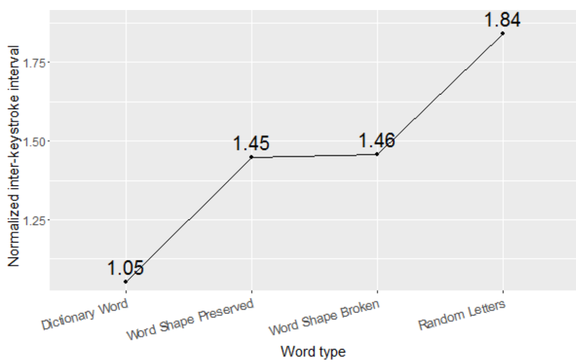
Table 3: Normalized mean and standard deviation of the number of mistakes made by participants on the "replay" condition on each of the first four attempts of every round.

	Attempt 1		Attempt 2		Attempt 3		Attempt 4	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
DictM	2.85	2.74	2.98	2.86	2.98	3.15	2.2	N/A
ShapeS	3.17	3.85	2.40	1.69	2.24	1.043	2.2	0.28
ShapeM	2.13	2.07	5.07	4.20	3.39	2.94	5	N/A
ShapeL	5.68	7.47	5.09	3.35	3.08	1.88	4.38	4.14
NoShapeS	2.46	2.18	2.69	1.54	2.65	2.07	4.65	5.93
NoShapeM	3.767	3.48	3.30	1.97	3.06	1.65	5.05	0.64
NoShapeL	6.19	6.63	4.10	2.34	5.17	3.10	6.44	5.24
RandS	1.91	2.06	1.99	1.90	1.87	1.46	2.18	1.78
RandM	4.94	3.94	4.03	2.13	5.48	1.90	4.37	4.20
RandL	4.72	4.89	2.71	1.81	4.03	3.52	6	2.83

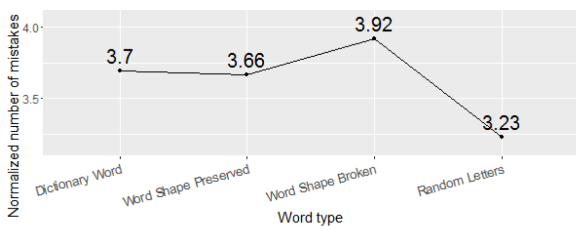
Our hypothesis **H2** expects there to be an improvement in accuracy as rounds are replayed. As before, we conducted a factorial ANOVA to examine the effects of Attempt and Round on the number of mistakes made. The results yielded a main effect for the round ($F(9, 284) = 3.2348, p < 0.001$), indicating that the typing accuracy of participants is significantly dependent on the round that was being played. The main effect of the attempt was not significant ($F(1, 284) = 0.0693, p > 0.1$). The interaction effect was also non-significant ($F(9, 284) = 0.4621, p > 0.1$). This results contradicts our hypothesis **H2**.

Familiarity with Words

For this analysis we look at how the different types of words in our game rounds affected speed and accuracy. In particular, we expected words that are more similar to real words to be typed faster (**H3**) and more accurately (**H4**). In decreasing order of similarity to real words we have dictionary words (*DictM*), dictionary words with transposed letters preserving the shape of the word (*ShapeS*, *ShapeM*, and *ShapeL*), dictionary words with transposed letters breaking the shape of the word (*NoShapeS*, *NoShapeM*, and *NoShapeL*), and random letters (*RandS*, *RandeM*, and *RandL*).



(a) Inter-keystroke interval



(b) Number of mistakes

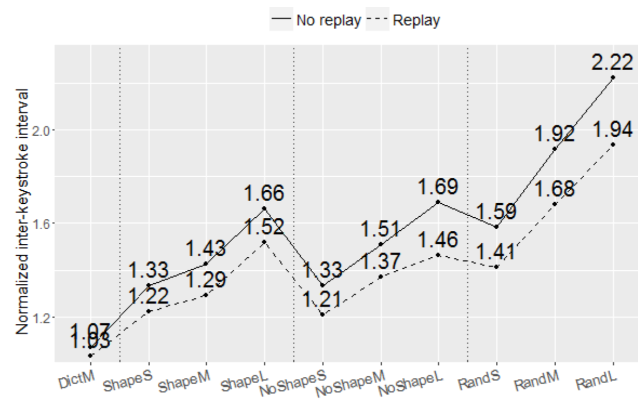
Figure 1: Comparison of the average normalized inter-keystroke interval and normalized number of mistakes by word type on the first attempt of every round.

The average IKI increases as the words participants typed resembled less real words (see Figure 1(a)), as predicted by **H3**. To evaluate significance of this difference we conducted a factorial ANOVA that explored the effects of word length, word type, and condition on IKI. The results yielded statistically significant interactions between the word type and word length ($F(4,28374) = 22.9631, p < 0.001$), between word length and condition ($F(2,28374) = 9.2675, p < 0.001$), and between word type and condition ($F(3,28374) = 10.4835, p < 0.001$). The interaction between word length, word type, and condition was not significant ($F(4,28374) = 0.1962, p > 0.1$). Simple main effects analysis showed significant differences in speed dependent on word length ($p < 0.001$), word type ($p < 0.001$), and condition ($p < 0.001$). This result is consistent with hypothesis **H3**.

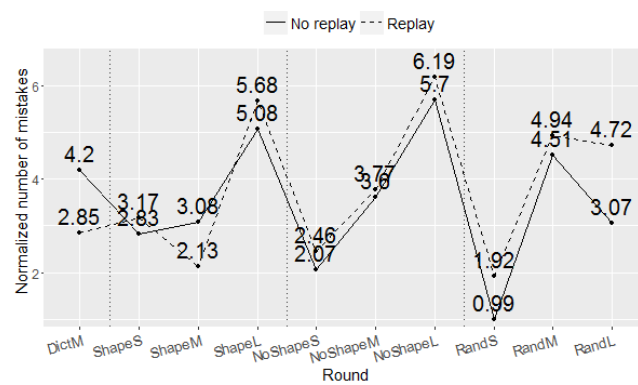
Our hypothesis **H4** expects participants to be more accurate on words that are closer to dictionary words. Figure 1(b) shows the number of mistakes made by our participants according to the type of word being typed. To evaluate these differences we conducted a factorial ANOVA that explored the effects of word length, word type, and condition on the number of mistakes made. The results yielded a statistically significant interaction between the word type and word length ($F(4,554) = 3.0836, p = 0.01578$). All the other interactions were not significant. Simple main effects analysis showed a significant difference in accuracy dependent on word length ($F(2,554) = 16.8025, p < 0.001$). We found no statistically significant difference in accuracy dependent on the type of the word. The lack of significance of the effect of the word type contradicts our hypothesis **H4**.

Additional Analyses

To obtain more insight we ran additional tests to compare the replay and no replay conditions on both speed an accuracy metrics, both on the first attempt of every round, and with up to 4 replays (for the replay condition; the no replay condition only had one attempt per round).



(a) Inter-keystroke interval



(b) Number of mistakes

Figure 2: Comparison of the average normalized inter-keystroke interval and normalized number of mistakes by condition on the first attempt of every round. The vertical lines separate rounds by word type.

When comparing the first attempt of every round between conditions we found that the mean IKIs of the replay condition were consistently smaller than those of the no replay condition (see Figure 2(a)). Using a Welch's independent-samples t-test, we found a significant difference in speed on the first attempt of every round between conditions ($t(18384) = 10.236, p < 0.001$).

In contrast, as shown on Figure 2(b), we don't see a clear distinction when comparing the number of mistakes made on the first attempt of every round between conditions. To determine significance difference we conducted a Welch's independent-samples t-test, which revealed no significant difference in accuracy on the first attempt of every round between conditions ($t(350.59) = -0.59744, p = 0.5506$).

Discussion

The above analysis confirms that typing speed improves with practice and when the words are more familiar. Surprisingly, we find that this improvement in speed is not accompanied by an improvement in typing accuracy neither with practice nor with familiarity with the words being typed. The number of mistakes made cannot be used to explain the reduction in IKI.

We saw that on the first attempt of the practice round, where the game experience is identical between conditions, all of our participants behave similarly. However, as the game progresses, participants in the replay condition significantly increase their typing speed without any improvement in the number of mistakes they make, indicating that the speed improvement is not attributable to an increase in accuracy. Because the only difference between conditions is the ability to replay rounds, a plausible explanation for this behavior lies in the fact that the cost (in terms of mathematical utility) of making mistakes is smaller than the reward of earning a higher compensation by typing faster, because the opportunity to replay the round is always there. This behavior is consistent with research on task accomplishment strategies, where there exists a trade-off between speed and accuracy (Barik, Chakraborty, Harrison, Roberts, & Amant, 2013; Gerjets, Scheiter, & Tack, 2000; Heitz, 2014). We find support for this explanation in our data when we compare the typing speed on the first attempt of every level between the replay and no replay conditions (see Figure 2(a)). On these first attempts, players in both conditions have had the same exposure to the words on each round, ruling out familiarity as an explanation for the significant difference in speed between conditions. We see that participants in the replay conditions are consistently and significantly faster than participants in the no replay condition after being exposed to the possibility of replaying, whereas this difference is non-existent on the first attempt of the practice round where they have not been exposed to this game mechanic.

Our results show that speed has a more direct relationship to the nature of what is being typed than the number of mistakes that are made while typing. This suggests that by inspecting typing speed a system can be more effective at detecting anomalies (and possibly identifying the cause of the anomaly) than looking at the number of incorrect attempts alone. Similarly, our results indicate that typing speed can be used to identify the familiarity to the text being typed, which can be used to compare to a known baseline.

Conclusions

In this work we explored how different cognitive processes affected typing patterns by manipulating the similarity of words to dictionary words, and by allowing participants to replay rounds of *The Typing Game*. We found that the typing speed improves with familiarity with words and with practice, but that these are independent of the number of mistakes that are made when typing. We also found that users exhibit different typing patterns when they are made aware of a penalty for mistakes than when they don't expect consequences for mistakes. Our results allow us to better understand the cognitive processes involved in typing.

There are several limitations to consider when interpreting our results. As mentioned earlier, we focused on establishing internal validity of our study, giving our first steps toward building cognitive models of input device interaction patterns. Firstly, because our sample was comprised mostly of Computer Science students, the typing proficiency of our participants is probably well above average, which is a threat to the external validity of our findings. Secondly, our game did not attempt to establish ecological validity, but was instead designed to elicit specific behaviors that manipulated the cognitive processes required to complete the game rounds. Thirdly, the nature of the words included in our game was also intentionally limited, and did not include numbers, uppercase letters, nor special characters. Despite these limitations, the empirical data we collected will allow us to generate cognitive models from interaction patterns of real users that can then be validated with a more representative sample and on multiple domains, pointing to avenues for future work.

Future Work

The data we collected from *The Typing Game* is incredibly rich, and this work presents preliminary results that we will use as stepping stones toward creating the cognitive models we discussed. We have already started working on a playback and visualization tool that will enable us to not only inspect and tag our data in more detail, but also to tweak and validate our assumptions as our models are created. We would like to explore how typing patterns differ with a more diverse character set, such as including capitalization, special characters and punctuation, and texts of different lengths (*e.g.,* a paragraph instead of a single word).

With a better understanding of typing phenomena and their relationship to cognition, we expect to validate our models on multiple domains. In particular, we aim to validate our models in domains that more closely resemble real-world tasks.

We also plan to investigate cognitive models of different input device usage. We expect that certain domains would benefit from models of multiple input devices simultaneously in order to improve the accuracy of their predictions, but also in order to provide a richer characterization of usage patterns.

Acknowledgments

This work was funded in part by the National Security Agency (grant number H98230-14-C-0139) and in part by the National Science Foundation (grant number IIS-1451172). The authors thank Brent Harrison and Arpan Chakraborty for their contributions to the design of *The Typing Game*.

References

Ahmed, A. A. E., & Traore, I. (2007). A new biometric technology based on mouse dynamics. *IEEE Transactions on Dependable and Secure Computing*, *4*(3), 165–179.

Barik, T., Chakraborty, A., Harrison, B., Roberts, D. L., & Amant, R. S. (2013). Speed/accuracy tradeoff in act-r models of the concentration game. In *Proceedings of the 2013 international conference on cognitive modeling* (pp. 281–286).

Bouwhuis, D., & Bouma, H. (1979). Visual word recognition of three-letter words as derived from the recognition of the constituent letters. *Perception & Psychophysics*, *25*(1), 12–22. doi: 10.3758/BF03206104

Card, S. K., Moran, T. P., & Newell, A. (1986). The model human processor: An engineering model of human performance. *Handbook of Perception and Human Performance*, *2*, 1–35.

Feit, A. M., Weir, D., & Oulasvirta, A. (2016). How we type: Movement strategies and performance in everyday typing. In *Proceedings of the 2016 chi conference on human factors in computing systems* (pp. 4262–4273). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/2858036.2858233` doi: 10.1145/2858036.2858233

Gerjets, P., Scheiter, K., & Tack, W. H. (2000). Resource-adaptive selection of strategies in learning from worked-out examples. In *Proceedings of the 22nd annual conference of the cognitive science society* (pp. 166–171). Erlbaum.

Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied*, *6*(4), 322–335.

Gray, W. D., & Fu, W. T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science*, *28*(3), 359–382.

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in neuroscience*, *8*(150). doi: 10.3389/fnins.2014.00150

John, B. E. (1996). Typist: A theory of performance in skilled typing. *Hum.-Comput. Interact.*, *11*(4), 321–355.

Monrose, F., & Rubin, A. (1997). Authentication via keystroke dynamics. In *Proceedings of the 4th acm conference on computer and comms security* (pp. 48–56).

Salthouse, T. A. (1986). Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological Bulletin*, *99*(3), 303–319.