Relationship between Dwell-Time and Model Human Processor for Dwell-based Image Selection

TOSHIYA ISOMOTO, University of Tsukuba, JAPAN
SHOTA YAMANAKA, Yahoo Japan Corporation, JAPAN
BUNTAROU SHIZUKI, University of Tsukuba, JAPAN

We investigated the relationship between dwell-time and the model human processor (MHP). First, we devised an equation that can represent the time taken for recognizing an image based on MHP. Then, we evaluated whether the equation can represent the time and wheter the time estimated by the equation matches the user's preferred dwell-time. The experiment consisted of two tasks: image selection with a button (button-task) and image selection with a dwell (dwell-task). From the results of the button-task, we found that the equation derived by MHP can estimate the time; the time taken for button selection was 662 ms on average, and the time estimated by the equation was 660 ms on average. Also, we showed that the estimated time represented the user's preferred dwell-time; all participants in the experiment answered that 500 ms and 600 ms were their preferred dwell-times.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; *HCI theory, concepts and models*; User studies.

Additional Key Words and Phrases: gaze-based interaction, cognitive system, user preference

ACM Reference Format:
Toshiya Isomoto, Shota Yamanaka, and Buntarou Shizuki. 2021. Relationship between Dwell-Time and Model Human Processor for Dwell-based Image Selection. In *ACM Symposium on Applied Perception 2021 (SAP '21), September 16–17, 2021, Virtual Event, France.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3474451.3476240

1 INTRODUCTION

In gaze-based interaction, looking has two significant purposes. One is for a cognitive process: a user recognizes the type and location of an object by looking at it. The other is for manipulation: a user selects an object (e.g., icon) or activates a command (e.g., copy) such as by left-clickikng with a mouse or pressing a hotkey.

The most-established gaze-based interaction is dwell selection, where the user selects an object by looking at it over a time-threshold (dwell-time) [Jacob 1990, 1991, 1993]. This time-threshold-based schema may cause the Midas-touch problem, which is a user's unwanted selection [Jacob 1990, 1991, 1993] because of the difficulty of distinguishing the purpose of the looking within the dwell-time. Researchers have attempted to solve the Midas-touch problem with a short dwell-time following the manner used in the human-computer interaction field, in which a faster and more accurate manipulation is superior. However, the dwell-time should not be too short because it should include a time for the cognitive process.

In this paper, we investigate the relationship between the dwell-time and the model human processor (hereinafter, MHP) [Card et al. 1983]. MHP is a well-known model that demonstrates a human's perceptual behavior in response to a visual (and auditory) stimulus. To examine the relationship between dwell-time and human behavior, we focus on MHP to explore the design of dwell selection instead of using the speed-accuracy trade-off. First, we devise an equation that can represent the time neccesary for recognizing a designated image based on MHP. Then, we evaluate whether the equation can represent the time and whether the time estimated by the equation fits a user's preferred dwell-time. The contribution of our paper will allow researchers and developers to use this user's preferred dwell-time to improve the performance of the dwell selection from the aspect of humans' perceptual behavior.

2 DWELL-TIME AND MODEL HUMAN PROCESSOR

First, we describe research that investigates the user's preferred dwell-time. Then, we describe the principle of MHP and how MHP demonstrates a human's perceptual behavior during dwell selection.

2.1 User Preferred Dwell-time

Researchers who desire to prevent the Midas-touch problem use 180–600 ms as the user's preferred dwell-time for dwell-typing [Majaranta et al. 2009; Mott et al. 2017; Pi and Shi 2017; Räihä and Ovaska 2012; Špakov and Miniotas 2004]. In a pilot study, 400 ms was chosen as a preferred dwell-time [Nayyar et al. 2017; Ware and Mikaelian 1987] for dwell selection of circles or images[1].

These dwell-times are determined from the speed-accuracy trade-off to achieve a faster selection and fewer Midas-touch problems. However, a user of a fast dwell selection with under 100 ms of dwell-time feels that the target is selected before the user looks at it [Isomoto et al. 2018]. Moreover, the visual search time varies with the size, arrangement, and color of a target or its content [Penkar et al. 2012; Sears et al. 2001]. As a dwell-time that can maintain the speed-accuracy trade-off of dwell selection, 650 ms is adequate for a low-cognitive task (search and select a two-digit); however, 1,100 ms is short for a high-cognitive task (search and select a two-word) [Zhang et al. 2011].

2.2 Model Human Processor

The MHP demonstrates a human's perceptual behavior in response to the visual (and auditory) stimulus by dividing the information-processing system into three subsystems: 1) the perception system, 2) the cognitive system, and 3) the motor system. Here, we describe the processes in these three subsystems responding to a visual stimulus. The perception system encodes a visual stimulus into visual code that the human can recognize and transmits it to working

[1] Unfortunately, a detailed experimental design and discussion were not provided.

SAP '21, September 16–17, 2021, Virtual Event, France

memory (WM); the time for the perception process (τ_p) is 100 [50–200] ms [Card et al. 1983]. With the visual code, the cognitive system recognizes, matches, classifies, and/or decides an action to the motor system to achieve the purpose; the time for the cognition process τ_c is 70 [25–170] ms [Card et al. 1983]. According to the decision from the cognitive system, the motor system requests an action such as pressing a button; the time for the motor process (τ_m) is 70 [30–100] ms [Card et al. 1983]. The ranges for each time indicate that a slowman (e.g., a novice) takes the maximum time and a fastman (e.g., an expert) takes a minimum time.

Using MHP, we can estimate the time necessary for a task. For a typing task of pressing the 'a' key, when the user looks at the stimulus, 1) the perception system transmits the visual code to the WM, requiring τ_p, the cognitive system 2-1) recognizes the visual code as a character (i.e., the visual code is 'a'), 2-2) matches the character and 'a' and 2-3) requests to press the key if they match, requiring $3\tau_c$, and 3) the motor system acts to press the button, requiring τ_m. The total time, according to MHP is

$$380\ [155\text{–}810] = \tau_\mathrm{p} + 3\tau_\mathrm{c} + \tau_\mathrm{m} = 100 + 3 \times 70 + 70.$$

Another example of a simple stimulus-response task where a user presses a specific key in reaction to a stimulus, the required processes for this task are 1) encoding the stimulus, 2) deciding, and 3) requesting to press the button. Therefore, the total time is

$$240\ [105\text{–}470] = \tau_\mathrm{p} + \tau_\mathrm{c} + \tau_\mathrm{m} = 100 + 70 + 70.$$

The processes for dwell-typing are similar to those of the typing task except for the process in the motor system. When a user looks at the 'a' key, 1) the perception system transmits the visual code, the cognitive system 2-1) recognizing the visual code as a character, 2-2) matching the recognized visual code and 'a', and 2-3) deciding to select the key (i.e., deciding to dwell at it) if they match. No motor action is required. Therefore, the total time is

$$310\ [125\text{–}710] = \tau_\mathrm{p} + 3\tau_\mathrm{c} = 100 + 3 \times 70.$$

Compared with previous research (e.g., [Majaranta et al. 2009; Räihä and Ovaska 2012], the range of the user's preferred dwell-time (180–600 ms) is similar to one calculated from the MHP (125–710 ms). Regarding the dwell selection with a shorter dwell time, if the dwell-time is under 100 ms, the target is selected before the information-processing system finishes the whole process, and thus, the user feels that the target is selected before the user looks at it. By using MHP, we can understand a dwell-time we should use for each task.

3 MODEL HUMAN PROCESSOR FOR DWELL-BASED IMAGE SELECTION

A dwell selection schema to select a designated image can be described as follows (in this case, the image is "bird'). When the user looks at an image, the information-processing system performs the following:

- **1)** transmit the visual code,
- **2-1)** recognize the visual code (the stimulus is 'wing'),
- **2-2)** classify the recognized code (this image which has wing is 'bird'),
- **2-3)** match the instruction and the visual code; if they match, the next step is 2-4); if they do not match, the process repeats from 2-1).
- **2-4)** decide whther to select the image. The process ends.

If the system selects a target before 2-4), the user may feel that the target is selected before recognizing the image. In our work, we assume a human usually looks at points within an image (or information in the real world) more than once to recognize the image and performs steps 2-1)–2-3) in one fixation. Note that the user's view specifies a certain point (i.e., the display coordinates) for gaze-based interaction. We use the MHP that describes the decision process, assuming that a user only considers one stimulus at a time. With the times τ_p, τ_c, and the number of fixations ($\mathrm{N_{fixation}}$) required for recognizing the image, the time that a user feels confortable with is estimated as

$$\tau_\mathrm{p} + (3\mathrm{N_{fixation}} + 1)\tau_\mathrm{c}. \tag{1}$$

Using this equation, if we predict $\mathrm{N_{fixation}}$, we can estimate the user's preferred dwell-time.

4 EXPERIMENT

We investigated the relationship between dwell-time and MHP by evaluating whether Equation 1 is appropriately derived from MHP and the equation can estimate the user's preferred dwell-time. This experiment consisted of two tasks: a button-task where users select a target by looking at it and then pressing a button, and a dwell-task where users select a target by performing dwell selection.

4.1 Participants and Apparatus

Sixteen volunteers (all Japanese males, our university students) aged 21 to 25 years (mean = 22.2) participated. Four had previously participated in an experiment using an eye tracker. None of the participants used an eye tracker daily.

We used a Tobii Eye Tracker 4C with a pro-license for research. A 24-inch (525×285 mm) display with a non-glare property for preventing reflections with a resolution of 2,560×1,440 pixels was used. Each participant placed the head approximately 60 cm from the display, and we positioned a keyboard used for controlling the task near the participant's hand. Before the task, the participants completed a 9-point calibration for the eye tracker with the supplied application. We used the same display for the button-task and dwell-task.

4.2 Task

Fig. 1 shows the display, which consists of instructions written in Japanese such as "Select (image)" where (image) represents a name among 28 kinds of image (e.g., "bird" or "horse") and the images[2]. We designed the arrangement and the size of the images allowing the robust selection against the performance of the eye-tracker. Regarding the arrangement, we made it a 5×4 matrix excluding the outermost area of the display, on the basis of findings that the eye-tracking performance decreases in the outermost area [Feit et al. 2017; Schuetz et al. 2020]. Each image was 40.0 mm of a square. Moreover, to exclude noize in the gaze data, we applied a low-pass

[2]we use the images from https://visualgenome.org/ which is licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/)

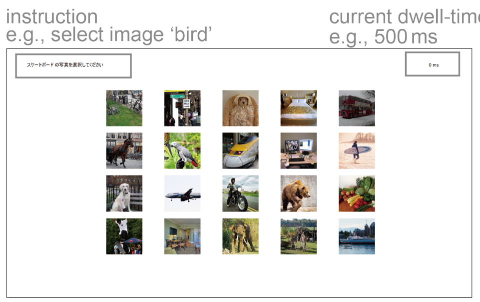

Fig. 1. Display used in the experiment.

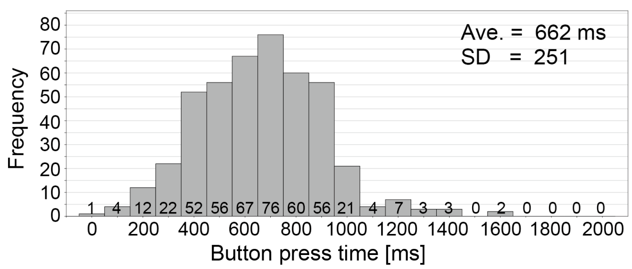

Fig. 2. Histogram of button press times for all participants.

filter expressed by $P_i = 0.25p_i + 0.75P_{i-1}$, where P_i is the i-th low-pass filtered gaze point and p_i is the i-th raw gaze point sampled from the eye tracker [Isomoto et al. 2020].

In each trial, we first displayed only the instruction and asked the participant to press the Space key to proceed. Then, we displayed 20 images (one designated image and 19 distractors). When the participant selected a designated image, we hid the 20 images and displayed the next instruction. We randomized the order of the instruction and randomly selected 19 distractors. During the tasks, to eliminate the effect of the Midas-touch problem on the user's preference, only the designated image was selectable, i.e., the other 19 distractors were not selectable. The participants could find the designated image freely by moving their gaze over the entire view. We displayed the current dwell-time during the dwell-task.

In the button-task, we asked participants to select 28 images in total. In the dwell-task, we investigated the dwell-time that participants wanted to use by examining 13 different dwell-times (0 ms, 100 ms, 200 ms, 300 ms, 400 ms, 500 ms, 600 ms, 700 ms, 800 ms, 900 ms, 1,000 ms, 1,500 ms, and 2,000 ms). We first asked participants to attempt the selection with each dwell-time and then answer the preference and comment about each dwell-time. After answering for each dwell-time, we asked the participants to answer the preferred dwell-time(s) that they want to use for dwell selection. Since all participants usually manipulate a computer with a mouse and do not manipulate them with gaze, we asked participants to start with the button-task that the participants are familiar with and then move on to the dwell-task. The average time required for the button-task was four minutes, and the dwell-task was 20 minutes per participant.

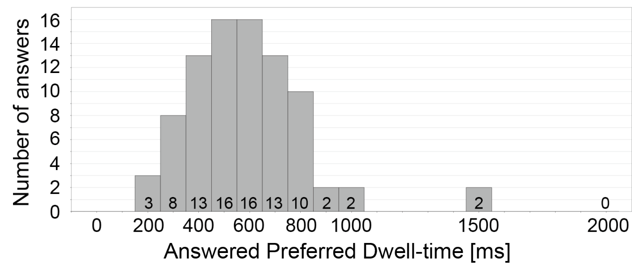

Fig. 3. Histogram of preferred dwell-times for all participants.

4.3 Results

We report the results of the button-task and dwell-task.

4.3.1 Button-task. First, we calculated the button press time from when the gaze entered the target until the participant pressed the button. We used the button press time as the ground-truth time that the participant requires to finish the cognitive process. The average button press time was 662 ms (SD=251). Histograms of the button press times in 100 ms are shown in Fig. 2. The distribution indicates that there was a peak at 700 ms.

4.3.2 Dwell-task. Regarding the comment for each dwell-time, when the dwell-time was shorter than 500 ms, the meanings of all of the responses were the same: "There was not enough time to understand the image, so I could not understand if the target had been selected correctly or not". Fifteen participants for 100 ms, ten participants for 200 ms, five participants for 300 ms, and three participants for 400 ms commented so. Similarly, when the dwell-time was longer than 500 ms, the meanings of all of the responses were the same: "I felt that the dwell-time was too long to select". Fig. 3 shows the user's preferred dwell-time. All participants (16/16) preferred 500 ms and 600 ms as a dwell-time and half of the participants (8/16) preferred 300–800 ms Fig. 3.

5 RELATIONSHIP BETWEEN RESULTS AND MODEL HUMAN PROCESSOR

We describe the relationship between the results and MHP.

5.1 Button Press Time

MHP explains the process that a participant recognizes a designated image and presses a button as follows: 1) encoding the stimulus into the visual code, 2-1) recognizing the visual code, 2-2) classifying the recognized code, 2-3) matching the classified code and instruction; if they match, the next step is 2-4); if they do not match, the next step is 2-1), 2-4) requesting to press a button, and 3) pushing the button. Given this process, the button press time estimated by MHP is expressed by Equation 2, which is the same as Equation 1 with the additional time in moter system.

$$\tau_\mathrm{p} + (3\mathrm{N}_\mathrm{fixation} + 1)\tau_\mathrm{c} + \tau_\mathrm{m} \quad (2)$$

The total time depends on how many times the process 2-1)–2-3) was repeated; if a participant can recognize that the designated image by one fixation (i.e., $\mathrm{N}_\mathrm{fixation}$= 1), the time is 450 [180–980] ms; by two fixations (i.e., $\mathrm{N}_\mathrm{fixation}$= 2), the time is 660 [255–1,490] ms. Note that because the size of the image used in this experiment was small, the I-VT (Velocity-Threshold Identification) algorithm [Salvucci and Goldberg 2000] with a velocity threshold of 30°/sec did not detect the saccade for 37.1% (166/448) of the attempts and detected one saccade for 59.1% (265/448) of the attempts that occur after the gaze enters the target; another 3.8% (17/448) of attempts include two or more saccades. We treated these saccades as outliers that did not affect our results, and thus did not add the saccade duration (i.e., the time taken for gaze movement) in this equation.

We counted the number of fixations representing points where a participant looks for recognizing the image. We used the I-DT (Dispersion-Threshold Identification) algorithm [Salvucci and Goldberg 2000] with a time window of 100 ms and dispersion threshold of 1.0°. As a result, the attempts containing one, two, three, four, and five or more fixations were 51 (11.4%), 253 (56.4%), 111 (24.8%), 26 (5.8%), and 7 (1.5%), respectively; the average number of fixations across whole attempts was 2.30 (SD=0.82). These numbers show that the participants could recognize the image used in this experiment as the designated one with one–three fixations. Using Equation 2, the estimated time and range with two fixations (660 [255–1,490] ms) cover 95.5% (428/448) of the button press times; and those with three fixations (870 [330–2,000] ms) cover 92.9% (416/448) of the button press times. This result shows that if we could predict the number of fixations required for recognizing the image, we can estimate the button press time with Equation 2.

5.2 Dwell-time

For the image selection with one to three fixations, the user's preferred dwell-times estimated from Equation 1 were

$$590\,\text{ms}\ [225\text{–}1,390] = 100 + (3 \times 2 + 1) \times 70\ (\text{N}_{\text{fixation}} = 2),$$

$$800\,\text{ms}\ [300\text{–}1,900] = 100 + (3 \times 3 + 1) \times 70\ (\text{N}_{\text{fixation}} = 3).$$

These times are consistent with the fact that the participants felt that a dwell-time shorter than 500 ms was too short to understand the image because the system selects the target during or before the cognitive process is completed. Since the cognitive process requires 225 ms even for the fastman, a period shorter than 200 ms is too short as a dwell-time for the image selection in our experiment. Moreover, all participants preferred 500 ms and 600 ms as dwell-time (Fig. 3) and the estimated user's preferred dwell-time from Equation 1 was 590 ms with two fixations that are caused 56.4% for the button-task. These results show that Equation 1 represents the time required for recognizing the designated image, and that we can estimate the user's preferred dwell-time using MHP.

6 LIMITATIONS

Since gaze movement heavily varies with the diversity of the participants, the diversity affects our finding concluded from all young male participants. If the preferred dwell-time is different among participants, we should determine a dwell-time for different user types, e.g., dwell-times for the young, for the elder, for males, and for females. Moreover, familiarity with the manipulation may also affect the findings. Therefore, further investigation is needed for revealing how diversity affects the user's preferred dwell-time.

The effect of visual feedback will affect the results, because previous research has reported that giving visual feedback to users may affect the usability [Majaranta et al. 2004, 2006]. In dwell-typing, visual feedback notifies that the user's gaze is inside a target or the duration regarding the dwell-time. Such visual feedback may affect the results of our experiment. For example, one participant responded, "because the dwell-time was so long, I was worried about whether my gaze was recognized correctly." for 2,000 ms of dwell-time. Thus, visual feedback emphasizing that the gaze is inside a target may positively affect the user's preferred dwell-time.

We used an image as the target in the experiment. However, in actual cases of manipulating a computer, some targets are a text such as a URL, in which case the information-processing would differ. According to a previous study [Rayner 1995], while reading a sentence, the distance of a saccade is 8 [1–15] characters, and the time until the next saccade (i.e., fixation time) is on average 250 (100–500) ms. This average time and range are similar to one estimated by the three cognitive processes (*recognition, classification, match*), i.e., 210 (75–510) ms. Moreover, the keystroke time range is 80–280 ms for the best typist and nonsecretary (not the slowman) [Card et al. 1980; Devoe 1967]. This knowledge may also be helpful to expand our findings to the text (i.e., words, phrases) selection.

Regarding $\text{N}_{\text{fixation}}$, predicting the number of fixations for a task is challenging. For the image selection, further investigation will help to determine the appropriate number. For text selection, as mentioned above, we can predict the number of saccades from the characters of the text; if the fixation and saccade occur alternatively, we can also predict the number of fixations.

7 CONCLUSION

We investigated the relationship between the dwell-time and the model human processor (MHP) [Card et al. 1983], unlike previous research which focuses on the relationship between dwell-time and the speed-accuracy trade-off. MHP demonstrates a human's perceptual behavior in response to a visual (and auditory) stimulus by dividing the information-processing system into three interacting subsystems: the perception system, cognitive system, and motors system. In addition, MHP defines the time requires for completing each system. We assumed that one fixation represents the process of *recognition, matching, and decision* in the cognitive system. Through this assumption and using MHP, we devised an equation that can estimate the time taken for recognizing a designated image with MHP. Using this equation, we then investigated the relationship between the dwell-time and MHP and whether the equation can estimate the user's preferred dwell-time. We conducted an image selection task with a button (button-task) and an image selection task with a dwell (dwell-task). From the results of the button-task, we found that the equation derived by MHP can estimate the dwell-time; the time taken for button selection was 662 ms on average, and the time estimated by the equation was 660 ms on average. In addition, we showed that this estimated time represented the user's preferred dwell-time; all participants in the experiment answered that 500 ms and 600 ms were their preferred dwell-time. From these results, we conclude that we can estimate a user's preferred dwell-time for image selection from MHP.

REFERENCES

Stuart K. Card, Thomas P. Moran, and Allen Newell. 1980. The Keystroke-Level Model for User Performance Time with Interactive Systems. *Communications of the ACM* 23, 7 (July 1980), 396–410. https://doi.org/10.1145/358886.358895

Stuart K. Card, Allen Newell, and Thomas P. Moran. 1983. *The Psychology of Human-Computer Interaction.* L. Erlbaum Associates Inc., USA.

Donald. B. Devoe. 1967. Alternatives to Handprinting in the Manual Entry of Data. *IEEE Transactions on Human Factors in Electronics* HFE-8, 1 (1967), 21–32. https://doi.org/10.1109/THFE.1967.233302

Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun Kane, and Meredith Ringel Morris. 2017. Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1118–1130. https://doi.org/10.1145/3025453.3025599

Toshiya Isomoto, Toshiyuki Ando, Buntarou Shizuki, and Shin Takahashi. 2018. Dwell Time Reduction Technique Using Fitts' Law for Gaze-Based Target Acquisition. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* (Warsaw, Poland) *(ETRA '18)*. Association for Computing Machinery, New York, NY, USA, 26:1–26:7. https://doi.org/10.1145/3204493.3204532

Toshiya Isomoto, Shota Yamanaka, and Buntarou Shizuki. 2020. Gaze-based Command Activation Technique Robust Against Unintentional Activation using Dwell-then-Gesture. In *Proceedings of Graphics Interface 2020* (University of Toronto) *(GI '20)*. Canadian Human-Computer Communications Society / Société canadienne du dialogue humain-machine, 256–266. https://doi.org/10.20380/GI2020.26

Robert J. K. Jacob. 1990. What You Look at is What You Get: Eye Movement-based Interaction Techniques. In *Proceedings of the 1990 CHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/97243.97246

Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-computer Interaction Techniques: What You Look at is What You Get. *ACM Transaction on Information Systems* 9, 2 (1991), 152–169.

Robert. J. K. Jacob. 1993. Eye Movement-Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces. *Advances in Human-Computer Interaction* 4 (1993), 151–190.

Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. 2009. Fast Gaze Typing with an Adjustable Dwell Time. In *Proceedings of the 2009 CHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 357–360. https://doi.org/10.1145/1518701.1518758

Päivi Majaranta, Anne Aula, and Kari-Jouko Räihä. 2004. Effects of Feedback on Eye Typing with a Short Dwell Time. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications* (San Antonio, Texas) *(ETRA '04)*. Association for Computing Machinery, New York, NY, USA, 139–146. https://doi.org/10.1145/968363.968390

Päivi Majaranta, I. Scott MacKenzie, Anne Aula, and Kari-Jouko Räihä. 2006. Effects of Feedback and Dwell Time on Eye Typing Speed and Accuracy. *Universal Access in the Information Society* 5, 2 (2006), 199–208. https://doi.org/10.1007/s10209-006-0034-z

Martez E. Mott, Shane Williams, Jacob O. Wobbrock, and Meredith Ringel Morris. 2017. Improving Dwell-Based Gaze Typing with Dynamic, Cascading Dwell Times. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2558–2570. https://doi.org/10.1145/3025453.3025517

Aanand Nayyar, Utkarsh Dwivedi, Karan Ahuja, Nitendra Rajput, Seema Nagar, and Kuntal Dey. 2017. OptiDwell: Intelligent Adjustment of Dwell Click Time. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus) *(IUI '17)*. Association for Computing Machinery, New York, NY, USA, 193–204. https://doi.org/10.1145/3025171.3025202

Abdul Moiz Penkar, Christof Lutteroth, and Gerald Weber. 2012. Designing for the Eye: Design Parameters for Dwell in Gaze Interaction. In *Proceedings of the 24th Australian Computer-Human Interaction Conference* (Melbourne, Australia) *(OzCHI '12)*. Association for Computing Machinery, New York, NY, USA, 479–488. https://doi.org/10.1145/2414536.2414609

Jimin. Pi and Bertram E. Shi. 2017. Probabilistic Ajustment of Dwell Time for Eye Typing. In *10th International Conference on Human System Interactions (HSI '17)*. IEEE, 251–257. https://doi.org/10.1109/HSI.2017.8005041

Kari-Jouko Räihä and Saila Ovaska. 2012. An Exploratory Study of Eye Typing Fundamentals: Dwell Time, Text Entry Rate, Errors, and Workload. In *Proceedings of the 2012 CHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 3001–3010. https://doi.org/10.1145/2207676.2208711

Keith Rayner. 1995. Eye Movements and Cognitive Processes in Reading, Visual Search, and Scene Perception. In *Eye Movement Research*, John M. Findlay, Robin Walker, and Robert W. Kentridge (Eds.). Studies in Visual Information Processing, Vol. 6. North-Holland, NLD, 3–22. https://doi.org/10.1016/S0926-907X(05)80003-0

Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (Palm Beach Gardens, Florida, USA) *(ETRA '00)*. Association for Computing Machinery, New York, NY, USA, 71–78. https://doi.org/10.1145/355017.355028

Immo Schuetz, T. Scott Murdison, and Marina Zannoli. 2020. A Psychophysics-Inspired Model of Gaze Selection Performance. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20)*. Association for Computing Machinery, New York, NY, USA, Article 25, 5 pages. https://doi.org/10.1145/3379156.3391336

Andrew Sears, Julie A. Jacko, Josey Chu, and Francisco Moro. 2001. The Role of Visual Search in the Design of Effective Soft Keyboards. *Behaviour & Information Technology* 20, 3 (2001), 159–166. https://doi.org/10.1080/01449290110049790 arXiv:https://doi.org/10.1080/01449290110049790

Oleg Špakov and Darius Miniotas. 2004. On-Line Adjustment of Dwell Time for Target Selection by Gaze. In *Proceedings of the 2004 Nordic Conference on Human-Computer Interaction (NordiCHI '04)*. Association for Computing Machinery, New York, NY, USA, 203–206. https://doi.org/10.1145/1028014.1028045

Colin Ware and Harutune H. Mikaelian. 1987. An Evaluation of an Eye Tracker As a Device for Computer Input. In *Proceedings of the 1987 CHI/GI Conference on Human Factors in Computing Systems and Graphics Interface* (Toronto, Ontario, Canada) *(CHI '87)*. Association for Computing Machinery, New York, NY, USA, 183–188. https://doi.org/10.1145/29933.275627

Xinyong Zhang, Pianpian Xu, Qing Zhang, and Hongbin Zha. 2011. Speed-Accuracy Trade-off in Dwell-Based Eye Pointing Tasks at Different Cognitive Levels. In *Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-Based Interaction* (Beijing, China) *(PETMEI '11)*. Association for Computing Machinery, New York, NY, USA, 37–42. https://doi.org/10.1145/2029956.2029967