

Rethinking the Dual Gaussian Distribution Model for Predicting Touch Accuracy in On-screen-start Pointing Tasks

SHOTA YAMANAKA, Yahoo Japan Corporation, Japan
HIROKI USUBA, Meiji University, Japan

The dual Gaussian distribution hypothesis has been used to predict the success rate of target pointing on touchscreens. Bi and Zhai evaluated their success-rate prediction model in *off-screen-start* pointing tasks. However, we found that their prediction model could also be used for on-screen-start pointing tasks. We discuss the reasons why and empirically validate our hypothesis in a series of four experiments with various target sizes and distances. The prediction accuracy of Bi and Zhai’s model was high in all of the experiments, with a 10-point absolute (or 14.9% relative) prediction error at worst. Also, we show that there is no clear benefit to integrating the target distance when predicting the endpoint variability and success rate.

CCS Concepts: • **Human-centered computing** → **HCI theory, concepts and models**; **Pointing**;

Keywords: Dual Gaussian distribution model; touchscreens; finger input; pointing

ACM Reference Format:

Shota Yamanaka and Hiroki Usuba. 2020. Rethinking the Dual Gaussian Distribution Model for Predicting Touch Accuracy in On-screen-start Pointing Tasks. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 4, ISS, Article 205 (November 2020). ACM, New York, NY. 20 pages. <https://doi.org/10.1145/3427333>

1 INTRODUCTION

Target acquisition is the most frequently performed operation on touchscreens. Tapping a small target, however, is sometimes an error-prone task, for reasons such as the “fat finger problem” [21, 48]. Hence, various techniques have been proposed to improve the precision of touch pointing [2, 48, 59]. Researchers have also sought to understand the fundamental principles of touch, e.g., touch-point distributions [4, 50]. As shown in these studies, the success rate in touch pointing is low, in particular for selecting a small target.

If touch GUI designers could compute the success rate of tapping a given target, they could determine target sizes that would strike a balance between usability and screen-space occupation. For example, suppose that a designer has to arrange many icons on a webpage. In this case, is a 5-mm diameter for each circular icon sufficiently large for accurate tapping? If not, then how about a 7-mm diameter? By how much can we expect the accuracy to be improved? Moreover, while larger icons can be more accurately tapped, they occupy more screen space. In that case, a webpage can be lengthened so that the larger icons fit, but this requires users to perform more scrolling operations to view and select icons at the bottom of the page. Another approach would be to split a page into several pages so that the target sizes remain. However, it is known that many users do not jump to the second or further pages (cf. the “Avoid the fold” and “Avoid the scroll” principles in web design [41]). Hence, designers have to carefully manage this tradeoff between

Authors’ addresses: Shota Yamanaka, Yahoo Japan Corporation, Tokyo, Japan; Hiroki Usuba, Meiji University, Tokyo, Japan.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2573-0142/2020/11-ART205 \$15.00

<https://doi.org/10.1145/3427333>

user performance and screen space, and they sometimes have to arrange targets that are smaller than common design guidelines (e.g., the Human Factors and Ergonomics Society recommends using >9.5-mm targets for touchscreens [27]).

Without a success-rate prediction model, designers have to conduct costly user studies to determine suitable target sizes on a webpage or app, but this strategy has low scalability. Accurate quantitative models would also be helpful for automatically generating user-friendly UIs [16, 40] and optimizing UIs [5, 14]. Furthermore, having such models would help researchers justify their experimental designs in evaluating novel interaction techniques that do not focus mainly on touch accuracy. For example, researchers could state, “According to the success-rate prediction model, 9-mm-diameter circular icons are assumed to be accurately (> 99%) selected by finger touching, and thus, this experiment is a fair one for comparing the usability of our proposed system and the baseline.”

To predict how successfully users will tap a target, Bi and Zhai proposed a model that computes the success rate solely from the target size (or width W) for both 1D and 2D pointing tasks [10]. They reasonably limited their model’s applicability to touch-pointing tasks *starting off-screen*, i.e., when a user’s finger moves from a position outside the touch screen. In this paper, we justify the use of the model for pointing with an on-screen start. After that, we empirically show, through a series of experiments, that the model has comparable prediction accuracy even for on-screen-start pointing. Our key contributions include:

- **Theoretical justification for applying Bi and Zhai’s success-rate prediction model to pointing tasks starting on-screen.** We found that the model is valid regardless of whether a pointing task starts on- or off-screen. This means that designers and researchers can predict success rates by using a single model.
- **Empirical verification of our hypothesis via four experiments.** We conducted 1D and 2D pointing experiments starting on-screen with randomized and controlled target distances (or amplitudes A). The results showed that we could predict the success rates with an absolute prediction error of ~10 points at worst, which corresponds to a 14.9% relative prediction error.

In short, the novelty of our study is that it extends the applicability of Bi and Zhai’s model to a variety of tasks, with support from empirical evidence. We thus expand the coverage of the model to other applications involving on-screen-start pointing, such as (a) tapping a “Like” button after scrolling through a social networking service feed, (b) successively inputting check marks on a questionnaire, and (c) typing on a software keyboard. Also, with this model, we can now predict the success rate for tapping a target with a given set of A and W values in Fitts’ law tasks on small touchscreens.

2 RELATED WORK

2.1 Success-Rate Prediction for Pointing Tasks

While a typical goal of pointing models is to predict the movement time MT , researchers have also tried to derive models to predict the success rate (or error rate). In particular, the model of Meyer et al. [37] is often cited as the first one to predict the error rate, but it does not account for the MT . In practice, the error rate increases as users move faster [63], and thus, Wobbrock et al. accounted for this effect in their model [54, 55]. Also, success-rate prediction for moving target selection is another hot topic in HCI [23, 24, 26, 42].

Simply speaking, when operators prioritize speed, the error rate increases. While Wobbrock et al. applied a time limit as an objective constraint by using a metronome in their study [54], this *speed-accuracy tradeoff* was also valid when the priority was subjectively biased [63]. Besides the case of rapidly aimed movements, the error rate has also been investigated for tapping on a static

button within a given temporal window [30–32]. Despite the recent importance of finger-touch operations on smartphones and tablets, however, the only model for predicting the success rate while accounting for finger-touch ambiguity is the work by Bi and Zhai on pointing from an off-screen start [10]. It would be useful if we could extend the validity of their model to other applications.

2.2 Improvements and Principles of Finger-Touch Accuracy

Examples of improving touch accuracy include using an offset from the finger contact point [11, 45, 48], dragging in a specific direction to confirm an intended target among many items [2, 39, 59], visualizing a contact point [53, 60], applying machine learning [51] or probabilistic modeling [9], and correcting hand tremor effects by using motion sensors [44]. In addition to these techniques, researchers have sought to understand why finger touch is less accurate compared with other input modalities such as a mouse cursor. One typical issue is the fat finger problem [21, 22, 48], in which a user wants to tap a small target, but the finger occludes it. Another issue is that finger touch has an unavoidable offset (spatial bias) from a user’s intended touch point to the actual touch position sensed by the system. Even if operators focus on accuracy by taking a sufficient length of time, the sensed touch point is biased from a small target [21, 22].

2.3 Success-Rate Prediction for Finger-Touch Pointing

2.3.1 Outline of Dual Gaussian Distribution Model. Previous studies have shown that the end-point distribution of finger touches follows a bivariate Gaussian distribution over a target [4, 19, 50]. Thus, a touch point **observed** by a system can be considered a random variable X_{obs} following a Gaussian distribution, $X_{\text{obs}} \sim N(\mu_{\text{obs}}, \sigma_{\text{obs}}^2)$, where μ_{obs} and σ_{obs} are the center and *SD* of the distribution, respectively. Bi, Li, and Zhai hypothesized that X_{obs} is the sum of two independent random variables consisting of **relative** and **absolute** components, both of which follow Gaussian distributions, $X_r \sim N(\mu_r, \sigma_r^2)$ and $X_a \sim N(\mu_a, \sigma_a^2)$ [8].

X_r is a relative component affected by the speed-accuracy tradeoff. When a user aims for a target more quickly, the relative endpoint variability σ_r increases. As indicated by Fitts’ law studies, if the acceptable endpoint tolerance W increases, then the user’s endpoint noise level σ_r also increases [12, 35].

X_a is an absolute component that reflects the precision of an input probe (a finger in this paper) and is independent of the task precision. Thus, even when a user taps a small target very carefully, there is still a spatial bias from the intended touch point [8, 21, 22]. The distribution of this bias is what σ_a models. Therefore, although σ_r can be reduced by a user aiming slowly at a target, σ_a cannot be controlled by placing such a priority on speed/accuracy. Note that the means of both components’ random variables (μ_r and μ_a) are assumed to tend to be close to the target center, $\mu_r \approx \mu_a \approx 0$, if the coordinate of the target center is defined as 0.

Again, the observed touch point is assumed to be a random variable that is the sum of two independent components:

$$X_{\text{obs}} = X_r + X_a \sim N(\mu_r + \mu_a, \sigma_r^2 + \sigma_a^2) \quad (1)$$

$\mu_{\text{obs}} (= \mu_r + \mu_a)$ is close to 0 on average, and σ_{obs}^2 is:

$$\sigma_{\text{obs}}^2 = \sigma_r^2 + \sigma_a^2 \quad (2)$$

When a user exactly utilizes the target size (W), $\sqrt{2\pi e}\sigma_r$ matches a given W (i.e., $4.133\sigma_r \approx W$) [9, 35]. Yet, users tend to be biased toward speed or accuracy, thus over- or underusing W [63]. Bi and Zhai assumed that using a fine probe of negligible size ($\sigma_a \approx 0$), such as a mouse cursor, makes

σ_r proportional to W . Thus, by introducing a constant α , we have:

$$\sigma_r^2 = \alpha W^2 \quad (3)$$

Then, replacing σ_r^2 in Equation 2 with Equation 3, we obtain:

$$\sigma_{\text{obs}}^2 = \alpha W^2 + \sigma_a^2 \quad (4)$$

Hence, by conducting a pointing task with several W values, we can run a linear regression on Equation 4 and obtain the constants α and σ_a . Accordingly, we can compute the endpoint variability for tapping a target of size W . We denote this endpoint variability computed from a regression expression as σ_{reg} :

$$\sigma_{\text{reg}} = \sqrt{\alpha W^2 + \sigma_a^2} \quad (5)$$

2.3.2 Revisiting Bi and Zhai's Studies on Success-Rate Prediction. Here, we revisit Bi and Zhai's first experiment on the *Bayesian Touch Criterion* [9]. They conducted a 2D pointing task with circular targets of diameter $W = 2, 4, 6, 8, \text{ and } 10$ mm. In their task, tapping the starting circle caused the first target to immediately appear at a random position. Subsequently, lifting the input finger off a target caused the next target to appear immediately. Hence, the participants successively tapped each new target as quickly and accurately as possible. The target distance was not predefined as A , unlike typical experiments involving Fitts' law. A possible way to analyze the effect of the movement amplitude would be to calculate A as the distance between the current target and the previous one; however, no such analysis was performed. Thus, even if the endpoint variability σ_{obs} was influenced by A , the effect was averaged.

By using Equation 5, the regression expressions of the σ_{reg} values on the x - and y -axes were calculated as:

$$\sigma_{\text{reg}_x} = \sqrt{0.0075W^2 + 1.6834} \quad \text{and} \quad \sigma_{\text{reg}_y} = \sqrt{0.0108W^2 + 1.3292} \quad (6)$$

Bi and Zhai then derived their success-rate prediction model [10]. Assuming a negligible correlation between the observed touch point values on the x - and y -axes (i.e., $\rho = 0$) gives the following probability density function for the bivariate Gaussian distribution:

$$P(x, y) = \frac{1}{2\pi\sigma_{\text{reg}_x}\sigma_{\text{reg}_y}} \exp\left(-\frac{x^2}{2\sigma_{\text{reg}_x}^2} - \frac{y^2}{2\sigma_{\text{reg}_y}^2}\right) \quad (7)$$

Then, the probability that the observed touch point falls within the target boundary D is:

$$P(D) = \iint_D \frac{1}{2\pi\sigma_{\text{reg}_x}\sigma_{\text{reg}_y}} \exp\left(-\frac{x^2}{2\sigma_{\text{reg}_x}^2} - \frac{y^2}{2\sigma_{\text{reg}_y}^2}\right) dx dy \quad (8)$$

where σ_{reg_x} and σ_{reg_y} are calculated from Equation 6. For a 1D horizontal bar target, whose boundary is defined as ranging from y_1 to y_2 , we can simplify the predicted probability for where the touch point Y falls on the target:

$$P(y_1 \leq Y \leq y_2) = \frac{1}{2} \left[\text{erf}\left(\frac{y_2}{\sigma_{\text{reg}_y}\sqrt{2}}\right) - \text{erf}\left(\frac{y_1}{\sigma_{\text{reg}_y}\sqrt{2}}\right) \right] \quad (9)$$

Note that the mean touch point μ of the probability density function is assumed to be ≈ 0 , thus eliminating it already from this equation. If the target width is W , then Equation 9 can be simplified further:

$$P\left(-\frac{W}{2} \leq Y \leq \frac{W}{2}\right) = \text{erf}\left(\frac{W}{2\sqrt{2}\sigma_{\text{reg}_y}}\right) \quad (10)$$

3 GENERALIZABILITY OF SUCCESS-RATE PREDICTION MODEL TO ON-SCREEN STARTING

Here, we discuss why Bi and Zhai’s model (Equations 8 and 10) can be applied to touch-pointing tasks starting on-screen, as well as possible concerns about this application. In their paper [9], Bi and Zhai stated, “We generalize the dual Gaussian distribution hypothesis from Fitts’ tasks—which are special target selection tasks involving both amplitude (A) and target width (W)—to the more general target-selection tasks which are predominantly characterized by W alone.” Therefore, to omit the effect of A when they later evaluated the success-rate prediction model, they told their participants to keep their dominant hands off the screen in natural positions and start from those positions in each trial [10], which simulated an off-screen-start pointing task.

The reason why Bi and Zhai limited the scope of their model to off-screen-start conditions was that they would limit the problem space to finger touching on mobile screens [62]. In most mobile interactions, users move their hand from off the screen and rest or get out of the eyesight to the screen path, and thus the off-screen-start condition was more appropriate. One the other hand, as mentioned by Zhai, the on-screen-start conditions are also common such as keyboard typing, which is the focus in this paper.

In generalizing the model to pointing tasks starting on-screen, one concern is the effect of A on the success rate. Even if we do not define the target distance A from the initial finger position, in actuality, the finger has an implicit travel distance, because “ A is less well-defined” [9] in off-screen-start tasks does not mean “there is no movement distance.” Thus, a pointing task predominantly characterized by W alone can also be interpreted as averaging the effects of A on touch-point distributions and success rates.

For example, suppose that a participant in off-screen-start tasks repeatedly taps a target 300 times. Let the implicit A value be 20 mm for 100 trials, 30 mm for another 100 trials, and 60 mm for the other 100. The success rates are independently calculated as (say) 95, 91, and 75%, respectively. If we do not distinguish the implicit A values, however, then the success rate would be $(95 + 91 + 75)/300 = 87\%$. This value is somewhat close to the 30-mm case (4 points; 4.60% difference from the prediction)¹, but the prediction errors for the other A values are more remarkable: 8 points (9.20%) and 12 points (13.8%) for 20 and 60 mm, respectively.

If the implicit or explicit movement distance A does not significantly change the success rate, such as from 88% for $A = 20$ mm to 86% for $A = 60$ mm, then we can use Bi and Zhai’s model regardless of whether pointing tasks start on- or off-screen. Now, the question is whether the success rate changes depending on the implicit or explicit A . According to the current prediction model (Equations 8 and 10), once W is given, the predicted success rate is determined by σ_{reg} . Hence, the debate revolves around whether the touch-point distribution is affected by the distance A . This is equivalent to asking whether Equation 4 ($\sigma_{\text{obs}}^2 = \alpha W^2 + \sigma_a^2$) is valid regardless of the value of A . In fact, the literature offers evidence on both sides of the question, as explained below.

Previous studies reported that the A does not strongly influence the endpoint distribution [8, 25, 63]. For these typical pointing tasks, participants perform closed-loop motions [20]. In contrast, the endpoint distribution for when participants perform a ballistic motion has been reported to be affected by A [20, 38, 46, 61]. Beggs et al. [6, 7] formulated the relationship in this way:

$$\sigma_{\text{obs}}^2 = \beta A^2 + \gamma \quad (11)$$

¹In this paper, we report both absolute and relative prediction errors. If the observed and predicted success rates are 91 and 87%, respectively, the absolute prediction error is reported to be 4 points, and the relative prediction error is $|91 - 87|/87 \times 100\% = 4.60\%$.

where σ_{obs} is valid for directions collinear and perpendicular to the movement direction, and β and γ are constants.

The critical threshold of whether participants perform a closed-loop or ballistic motion depends on Fitts' original index of difficulty, $ID = \log_2(2A/W)$. When ID is less than 3 or 4 bits, a pointing task can be accomplished with only a ballistic motion [17, 20]. While the critical ID changes depending on the experimental conditions, an extremely easy task with a short A or large W generally does not require precise closed-loop operations. Hence, we theoretically assume that the endpoint distribution σ_{obs} and the success rate change depending on A .

Nevertheless, the changes in σ_{obs} and the success rate due to A might be small in practice. This is because extremely short A values, where users can accomplish a pointing task with only a single ballistic movement [17, 20], also reduce the impact on σ_{obs} according to Equation 11. If so, then, from a practical viewpoint, it would not be problematic to apply Bi and Zhai's success-rate prediction model to pointing from an on-screen start; therefore, we empirically validated this.

4 EXPERIMENTS

We ran experiments involving 1D and 2D pointing tasks. For each dimensionality, we conducted (a) successive pointing tasks in which a target appeared at a random position immediately after the previous target was tapped and (b) discrete pointing tasks in which the target distance A was predefined. Under condition (a), we disregarded the effect of movement distance; this was a fair modification of Bi and Zhai's success-rate prediction experiments [10] to an on-screen-start condition. Under condition (b), we separately predicted the success rates for each value of A to empirically evaluate the effect of movement distance on the prediction accuracy. We thus conducted four experiments composed of 1D and 2D target conditions:

Exp. 1. Successive 1D pointing task: horizontal bar targets appeared at random positions.

Exp. 2. Discrete 1D pointing task: a start bar and a target bar were displayed with distance A between them.

Exp. 3. Successive 2D pointing task: circular targets appeared at random positions.

Exp. 4. Discrete 2D pointing task: a start circle and a target circle were displayed with distance A between them.

Experiments 1 and 2 were conducted on the first day and performed by the same 12 participants. Although we explicitly labeled these as Experiments 1 and 2, their order was balanced among the 12 participants. Similarly, on the second day, 12 participants were divided into two groups, and the order of Experiments 3 and 4 was balanced. Each set of two experiments took less than 40 min per participant.

We used an iPhone XS Max (2.5-GHz CPU; 4-GB RAM; iOS 12; 1242 × 2688 pixels, 6.5-inch display, 458 ppi; 208 g). The experimental system was implemented with JavaScript, HTML, and CSS. The web page was viewed with the Safari app. After eliminating the navigation-bar areas, the canvas resolution was 414 × 719 pixels, giving 5.978 pixels/mm. The system was set to run at 60 fps. We used the takeoff positions as tap points, as in [8–10, 57, 58].

The participants were asked to sit in an office chair in a silent room. As shown in Figure 1a, each participant held the smartphone with the nondominant (left) hand and tapped the screen with the dominant (right) hand's index finger. They were instructed not to rest their hands or elbows on their laps.

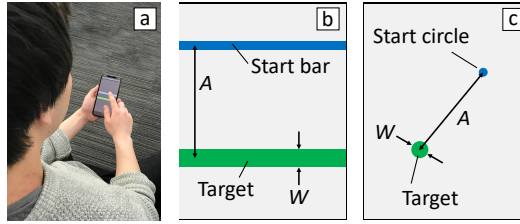


Fig. 1. (a) Experimental environments, and the visual stimuli used in (b) Experiment 2 and (c) Experiment 4.

5 EXPERIMENT 1: 1D TASK WITH RANDOM AMPLITUDES

5.1 Participants

Twelve university students joined in this study (2 female, 10 male; 20 to 25 years, $M = 23.0$, $SD = 1.41$). They were right-handed and daily smartphone users. Five participants used iOS smartphones daily, and seven used Android smartphones. They each received 45 USD for performing Experiments 1 and 2.

5.2 Task and Design

A 6-mm-high start bar was initially displayed at a random position on the screen. When a participant tapped it, the first target bar immediately appeared at a random position. The participant was tasked with successively tapping new targets that appeared upon lifting the finger off. If a target was missed, a beep sounded, and the participant had to re-aim for the target. If the participant succeeded, a bell sounded. To reduce the negative effect of screen edges, the random target position was at least 11 mm away from the top and bottom screen edges [3].

This task had a single-factor within-subjects design with an independent variable of W : 2, 4, 6, 8, and 10 mm. Because the target distance from the previous target changed randomly, this task required from ballistic to closed-loop motions. The dependent variables were the observed touch-point distribution on the y-axis, σ_{obs_y} , and the success rate. The touch-point bias was measured from the target center with a sign [56]. First, the participants performed 20 trials as practice, which included 4 repetitions of the 5 W values appearing in random order. In each *session*, the W values appeared 10 times in random order. The participants were instructed to successively tap the target as quickly and accurately as possible in a single session. They each completed four sessions as data-collection trials. In total, we recorded $5_W \times 10_{\text{repetitions}} \times 4_{\text{sessions}} \times 12_{\text{participants}} = 2400$ trials.

5.3 Results

We removed 13 outlier trials (0.54%) that had tap points at least 15 mm from the target center [9]². According to observation, such outliers resulted mainly from participants accidentally touching the screen with the thumb or little finger. For consistency with Bi and Zhai's work [10], we computed the regression between $\sigma_{\text{obs}_y}^2$ and W^2 to validate Equation 4, compared the observed and computed touch-point distributions (σ_{obs_y} and σ_{reg_y} , respectively), and compared the observed and predicted success rates. Although our results showed that the dependent variables did not pass the Shapiro-Wilk test ($\alpha = 0.05$) in some cases, it is known that ANOVA is robust against violations of the normality test assumptions [13, 36], and thus, we consistently ran repeated-measures ANOVAs.

²Using a fixed distance may affect W levels differently. For example, more outliers could be observed for $W = 10$ mm than $W = 2$ mm. Yet, we here maintain consistency with the previous study [9].

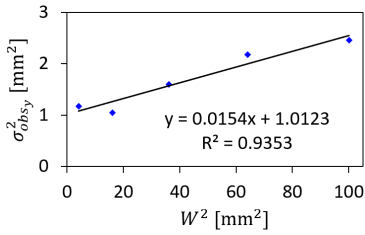


Fig. 2. $\sigma_{obs,y}^2$ vs. W^2 in Experiment 1.

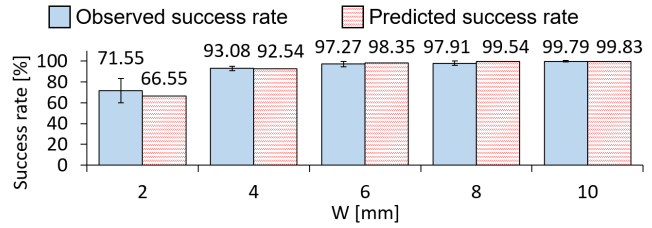


Fig. 3. Observed vs. predicted success rates in Experiment 1. Error bars show *SDs* across participants throughout this paper.

5.3.1 Touch-Point Distribution. The W had a significant main effect on $\sigma_{obs,y}$ ($F_{4,44} = 11.18$, $p < 0.001$, $\eta_p^2 = 0.50$). Shapiro-Wilk tests showed that the touch points followed normal distributions under 47 of the 60 conditions ($= 5_W \times 12_{participants}$), or 78.3%. Figure 2 shows the regression expression for $\sigma_{obs,y}^2$ versus W^2 for validating Equation 4. The assumption of a linear relationship for these variances even for touch-pointing operations with an on-screen start was supported with $R^2 = 0.9353$. The differences between the computed $\sigma_{reg,y}$ and observed $\sigma_{obs,y}$ values were less than 0.1 mm (< 1 pixel), as obtained by taking the square roots of the vertical distances between the points and the regression line in Figure 2.

5.3.2 Success Rate. Among the 2387 ($= 2400 - 13$) non-outlier data points, the participants successfully tapped the target in 2194 trials, or 91.91%. As shown by the blue bars in Figure 3, the observed success rate increased from 71.55 to 99.79% with the increase in W , which had a significant main effect ($F_{4,44} = 58.37$, $p < 0.001$, $\eta_p^2 = 0.84$).

By applying the regression expression $\sigma_{reg,y} = \sqrt{0.0154W^2 + 1.0123}$ (from Figure 2) to Equation 10, we computed the predicted success rates for each W , as represented by the red bars in Figure 3. The difference from the observed success rate was 5.00 points (7.51%) at most. These results show that we could predict the success rate solely from the target size W , with a mean absolute error *MAE* of 1.657% for $N = 5$ data points. This indicates the applicability of Bi and Zhai’s model with our on-screen start condition.

6 EXPERIMENT 2: 1D TASK WITH PRESET AMPLITUDES

6.1 Task and Design

Figure 1b shows the visual stimulus used in Experiment 2. At the beginning of each trial, a 6-mm-high blue start bar and a W -mm-high green target bar were displayed at random positions with distance A between them. When a participant tapped the start bar, it disappeared, and a click sounded. Then, if the participant successfully tapped the target, a bell sounded, and the next start and target bars appeared. If the participant missed the target, s/he had to aim at it until successfully tapping it. In this case, the trial was not restarted from tapping the start bar. The participants were instructed to tap the target as quickly and accurately as possible after tapping the start bar.

This study was a 4×5 within-subjects design. We included four A s (20, 30, 45, and 60 mm) and five W s (2, 4, 6, 8, and 10 mm). This covered a wide range of ID s requiring ballistic to closed-loop motions. Each $A \times W$ combination was used for 16 repetitions, following a single repetition of practice trials. In total, we recorded $4_A \times 5_W \times 16_{repetitions} \times 12_{participants} = 3840$ trials.

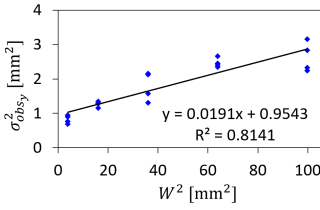


Fig. 4. $\sigma_{\text{obs}_y}^2$ vs. W^2 in Experiment 2.

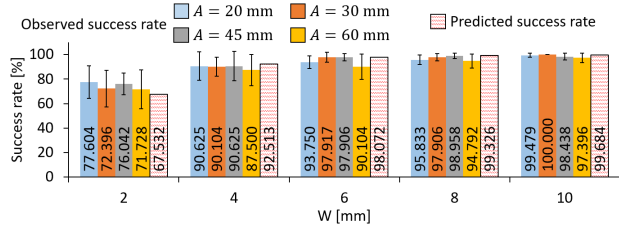


Fig. 5. Observed vs. predicted success rates in Experiment 2.

6.2 Results

6.2.1 Touch-Point Distribution. We removed four outlier trials (0.10%) that had tap points at least 15 mm from the target center. We found significant main effects of A ($F_{3,33} = 2.949$, $p < 0.05$, $\eta_p^2 = 0.21$) and W ($F_{4,44} = 72.63$, $p < 0.001$, $\eta_p^2 = 0.87$) on σ_{obs_y} , but no significant interaction of $A \times W$ ($F_{12,132} = 1.371$, $p = 0.187$, $\eta_p^2 = 0.11$). We found that 218 of the 240 conditions ($4_A \times 5_W \times 12_{\text{participants}}$) passed the normality test, or 90.8%.

Figure 4 shows the regression expression for $\sigma_{\text{obs}_y}^2$ versus W^2 , with $R^2 = 0.8141$ for $4_A \times 5_W = 20$ data points. When we merged the four $\sigma_{\text{obs}_y}^2$ values for each A as we did with the amplitudes in Experiment 1, we obtained five data points with $R^2 = 0.9171$ (the regression constants did not change). Using the regression expression $\sigma_{\text{reg}_y} = \sqrt{0.0191W^2 + 0.9543}$ (from Figure 4), we computed σ_{reg_y} for each W . The differences between the σ_{reg_y} and σ_{obs_y} values were less than 0.2 mm (~ 1 pixel). As a check, the average σ_{obs_y} values for $A = 20, 30, 45,$ and 60 mm were 1.345, 1.279, 1.319, and 1.415 mm, respectively, giving a difference of 0.136 mm at most.

The conclusion of the small effect of A was supported by a repeated-measures ANOVA; for σ_{obs_y} , a pairwise test with Bonferroni correction as the p -value adjustment method showed only one pair having a significant difference between $A = 45$ and 60 mm ($p < 0.05$; $|1.319 - 1.415| = 0.096$ mm < 1 pixel). These results indicate that we could compute the touch-point distributions regardless of A at a certain degree of accuracy in most cases.

6.2.2 Success Rate. Among the 3836 ($= 3840 - 4$) non-outlier data points, the participants successfully tapped the target in 3489 trials, or 90.95%. We found significant main effects of A ($F_{3,33} = 4.124$, $p < 0.05$, $\eta_p^2 = 0.27$) and W ($F_{4,44} = 45.03$, $p < 0.001$, $\eta_p^2 = 0.80$) on the success rate, but no significant interaction of $A \times W$ ($F_{12,132} = 0.681$, $p = 0.767$, $\eta_p^2 = 0.058$). Figure 5 shows the observed and predicted success rates. The largest difference was $77.60 - 67.53 = 10.07$ points (14.9%) under the condition of $A = 20$ mm $\times W = 2$ mm. This is comparable to Bi and Zhai's success-rate prediction [10], in which the largest difference (9.74 points; 14.2%) was observed for $W = 2.4$ mm for a 1D vertical bar target. In Experiment 2, the MAE was 3.266% for $N = 20$ data points.

7 EXPERIMENT 3: 2D TASK WITH RANDOM AMPLITUDES

The experimental designs were almost entirely the same as in Experiments 1 and 2, except that circular targets were used in Experiments 3 and 4. Here, the target size W means the circle's diameter. The random target positions were set at least 11 mm from the edges of the screen. For Experiment 3, we used the same task design as in Experiment 1: $5_W \times 40_{\text{repetitions}} \times 12_{\text{participants}} = 2400$ data points.

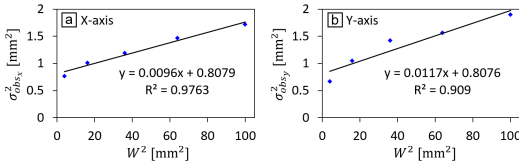


Fig. 6. W^2 vs. (a) $\sigma_{\text{obs}_x}^2$ and (b) $\sigma_{\text{obs}_y}^2$ in Experiment 3.

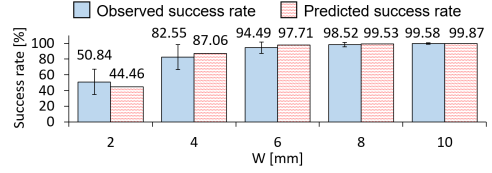


Fig. 7. Observed vs. predicted success rates in Experiment 3.

7.1 Participants

Twelve university students joined in this study (3 female, 9 male; 19 to 25 years, $M = 22.2$, $SD = 2.12$), which included 9 new participants. All were right-handed and daily smartphone users. Nine of the participants used iOS smartphones daily, and three used Android. They each received 45 USD for performing Experiments 3 and 4.

7.2 Results

7.2.1 Touch-Point Distribution. We removed 33 outlier trials (1.375%) that had tap points at least 15 mm from the target center. The W had significant main effects on σ_{obs_x} ($F_{4,44} = 15.96$, $p < 0.001$, $\eta_p^2 = 0.59$) and σ_{obs_y} ($F_{4,44} = 25.71$, $p < 0.001$, $\eta_p^2 = 0.70$). The touch points on the x - and y -axes followed normal distributions for 55 (91.7%) and 53 (88.3%) out of 60 conditions, respectively. Under 41 (68.3%) conditions, the touch points followed bivariate normal distributions. Figure 6 shows the regression expressions for $\sigma_{\text{obs}_x}^2$ and $\sigma_{\text{obs}_y}^2$ versus W^2 . Using these regressions ($\sigma_{\text{reg}_x} = \sqrt{0.0096W^2 + 0.8079}$ and $\sigma_{\text{reg}_y} = \sqrt{0.0117W^2 + 0.8076}$), we computed the touch-point distributions for each W . The differences between the computed σ_{reg} and observed σ_{obs} values were at most 0.05 and 0.2 mm for the x - and y -axes, respectively.

7.2.2 Success Rate. Among the 2367 (= 2400 – 33) non-outlier data points, the participants successfully tapped the target in 2017 trials, or 85.21%. As shown by the blue bars in Figure 7, the observed success rate increased from 50.84 to 99.58% with W , which had a significant main effect ($F_{4,44} = 59.24$, $p < 0.001$, $\eta_p^2 = 0.84$). As represented by the red bars, we computed the predicted success rates for each W by applying the regression expressions for σ_{reg_x} and σ_{reg_y} to Equation 8. The differences from the observed success rates were less than 7 points (14.3% at most). These results show that we could predict the success rate from the W , with $MAE = 3.082\%$ for $N = 5$ data points.

8 EXPERIMENT 4: 2D TASK WITH PRESET AMPLITUDES

We used the same task design as in Experiment 2: $4_A \times 5_W \times 16_{\text{repetitions}} \times 12_{\text{participants}} = 3840$ data points. Figure 1c shows the visual stimulus.

8.1 Results

8.1.1 Touch-Point Distribution. We removed nine outlier trials (0.23%) that had tap points at least 15 mm from the target center. For σ_{obs_x} , we found a significant main effect of W ($F_{4,44} = 24.12$, $p < 0.001$, $\eta_p^2 = 0.69$), but not of A ($F_{3,33} = 0.321$, $p = 0.810$, $\eta_p^2 = 0.028$). No significant interaction of $A \times W$ was found ($F_{12,132} = 0.950$, $p = 0.500$, $\eta_p^2 = 0.079$). For σ_{obs_y} , we found significant main effects of A ($F_{3,33} = 3.833$, $p < 0.05$, $\eta_p^2 = 0.26$) and W ($F_{4,44} = 48.35$, $p < 0.001$, $\eta_p^2 = 0.82$), but no significant interaction of $A \times W$ ($F_{12,132} = 1.662$, $p = 0.082$, $\eta_p^2 = 0.13$). Pairwise comparisons showed that only $A = 30$ and 60 mm had a significant difference ($p < 0.05$). The touch points on the x - and y -axes

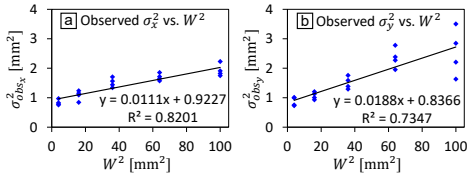


Fig. 8. W^2 vs. (a) $\sigma_{\text{obs}_x}^2$ and (b) $\sigma_{\text{obs}_y}^2$ in Experiment 4.

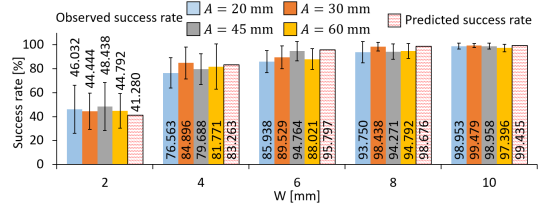


Fig. 9. Observed vs. predicted success rates in Experiment 4.

followed normal distributions for 224 (93.3%) and 218 (90.8%) of the 240 conditions, respectively. Under 184 (76.7%) conditions, the touch points followed bivariate normal distributions.

Figure 8 shows the regression expressions for $\sigma_{\text{obs}_x}^2$ and $\sigma_{\text{obs}_y}^2$ versus W^2 , with $R^2 = 0.8201$ and 0.7347, respectively, for $N = 20$ data points. When we merged the four $\sigma_{\text{obs}_x}^2$ and $\sigma_{\text{obs}_y}^2$ values for each A , we obtained $N = 5$ data points with $R^2 = 0.9137$ and 0.9425, respectively (the regression constants did not change). Using these regressions ($\sigma_{\text{reg}_x} = \sqrt{0.0111W^2 + 0.9227}$ and $\sigma_{\text{reg}_y} = \sqrt{0.0188W^2 + 0.8366}$), we computed the touch-point distributions under each condition of $A \times W$. The differences between the computed σ_{reg} and observed σ_{obs} values were less than 0.2 mm on the x -axis and less than 0.5 mm on the y -axis.

8.1.2 Success Rate. Among the remaining 3831 ($= 3840 - 9$) non-outlier data points, 3145 trials were successful (82.09%). We found a significant main effect of W ($F_{4,44} = 120.0$, $p < 0.001$, $\eta_p^2 = 0.92$), but not of A ($F_{3,33} = 2.100$, $p = 0.119$, $\eta_p^2 = 0.16$). The interaction of $A \times W$ was not significant ($F_{12,132} = 0.960$, $p = 0.490$, $\eta_p^2 = 0.080$). Figure 9 shows the observed and predicted success rates. The largest difference was $95.80 - 85.94 = 9.86$ points (11.5%) for $A = 20$ mm and $W = 6$ mm. The MAE was 3.671% for $N = 20$ data points.

9 DISCUSSION

9.1 Accuracy in Predicting Success Rates

Throughout the experiments, the prediction errors were about as low as in Bi and Zhai's pointing tasks with an off-screen start [10]: 10.07 points (14.9%) at most in our case (for $A = 20$ mm \times $W = 2$ mm in Experiment 2), versus 9.74 points (14.2%) at most in Bi and Zhai's case (2.4 mm, vertical bar target). As in their study, we found that the success rate approached 100% as W increased, and thus, the prediction errors tended to become smaller. Therefore, the model accuracy should be judged from the prediction errors for small targets.

The largest prediction error in Bi and Zhai [10] was under the condition of $W = 2.4$ mm. In comparison, the largest prediction error in our experiments was under a condition with slightly smaller targets of $W = 2$ mm. Also, while Bi and Zhai checked the prediction errors under nine conditions in total [10] (three W values for three target shapes), we checked the prediction errors under $5 + 20 + 5 + 20 = 50$ conditions (Experiments 1 to 4, respectively). Thus, these differences may have given more chances to show a higher prediction error in our tasks. In addition, although we used 2 mm as the smallest W , such a small target is not often used in practical touch UIs. Therefore, the slightly larger prediction error in our results should be less critical in actual usage.

We observed that the error rates in our experiments were sometimes higher than those in typical pointing studies. For example, the maximum error rate was 56% (Experiment 4), but it is assumed that the error rate should be close to 4% [35, 47]. However, as Gori et al. recently identified, this

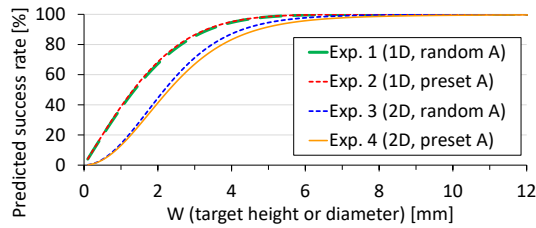


Fig. 10. Predicted success rate with respect to the W .

4% criterion is arbitrary, and error rates can change depending on the target size [18], which is supported by existing empirical data [10, 15, 49, 54]. It is not unusual for the error rate to be quite high, in particular for touch pointing tasks, e.g., an error rate of approximately 20% on average for $W = 3, 5,$ and 7-mm diameter circular targets [9] and 64% for 2.4-mm circles [10]. If we analyze only the two largest target sizes ($W = 8$ and 10-mm conditions), the maximum error rate was 6.25% in Experiment 4 (Figure 9, $W = 8$ mm, $A = 20$ -mm condition), which is not remarkably high compared with previous studies. Therefore, the error rate is expected to increase as the target size decreases, and thus, the chance for the prediction error to increase also increases.

We also found that our concern that the prediction accuracy might drop, depending on the A values, was not a critical issue as compared with tasks using an off-screen start [10]. Hence, the comparable prediction accuracy observed in our experiments empirically shows that Bi and Zhai’s model can be applied to pointing tasks with an on-screen start, regardless of whether the effect of A is averaged (Experiments 1 and 3) or not (2 and 4).

Figure 10 plots the predicted success rate with respect to W , which can help designers choose the appropriate size for a GUI item. This also provides evidence that conducting costly user studies to measure success rates for multiple W values has low scalability. For example, from the data in Experiment 4, the success rates sharply rose from $W = 1$ to 6 mm. Hence, for example, even if the error rate is measured for $W = 2, 6,$ and 10 mm, it would be difficult to accurately predict error rates for other W values such as 3 mm. Therefore, without an appropriate success-rate prediction model, designers have to conduct user studies with fine-grained W values, e.g., 1 to 10 mm at 1-mm intervals. However, 1-mm intervals are still not sufficient; the predicted success rate “jumps up” from 41.3 to 67.0% for $W = 2$ and 3 mm, respectively. Therefore, this strategy should be considered inappropriate due to the scalability issue, particularly for small targets.

9.2 Adequacy of Experiments

In our experiments, the endpoint distributions were not normal in some cases. Those results did not pass the assumption of a dual Gaussian distribution model. Figure 11 allows us to visually check the distributions. Some conditions did not exhibit normal distributions, e.g., Figure 11e showing a dent at the peak. This could be partly due to the small numbers of trials in our experiments. Still, according to the central limit theorem, it is reasonable to assume that the distributions should approach normal after a sufficient number of trials.

We also checked the Fitts’ law fitness. Using the Shannon formulation [35], we found that the error-free MT data showed excellent fits³ for Experiments 2 and 4, respectively, by using $N = 20$ data points; $MT = 132.0 + 90.29 \times \log_2(A/W + 1)$ with $R^2 = 0.9807$ and $MT = 114.3 + 97.91 \times \log_2(A/W + 1)$ with $R^2 = 0.9900$. The indexes of performance, $IP (= 1/slope)$, were 11.08 and 10.21

³Results for the effective width method [12, 35] and FFitts law [8] obtained by taking failure trials into account were also analyzed. Because of space limitations, we decided to focus on success-rate prediction in this paper.

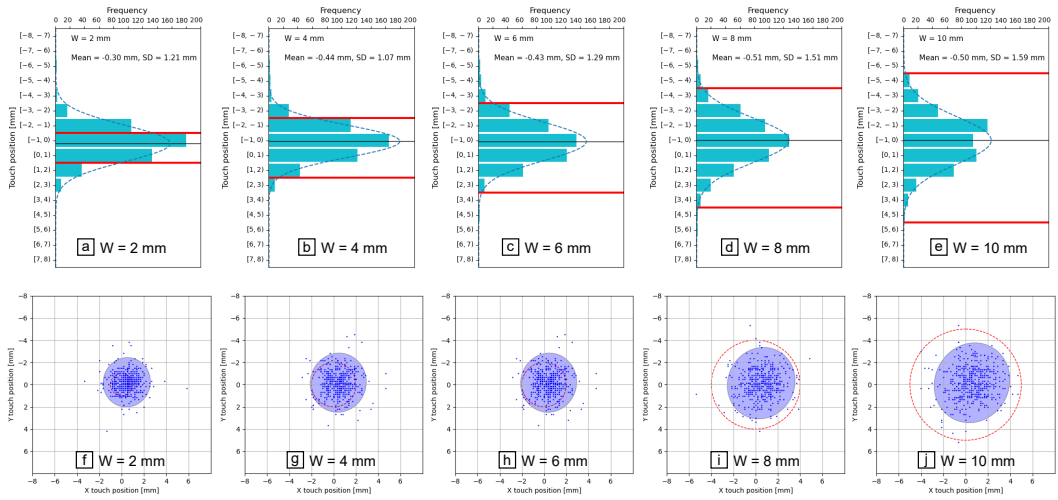


Fig. 11. Histograms and 95% confidence ellipses using the all non-outlier data in Experiments (a–e) 1 and (f–j) 3. For 1D tasks, the histograms show the frequencies of tap positions, the dashed curved lines show the normal distributions using the mean and σ_{obs_y} data, the two red bars are the borderlines of the target, and the black bar shows the mean of tap positions. For 2D tasks, the blues dots are tap positions, the light blue ellipses are 95% confidence ellipses of tap positions, and the red dashed circles are target areas. For all tasks, the 0-mm positions on the x- and y-axes were aligned to the centers of targets.

bits/s, close to those in Pedersen and Hornbæk’s report on error-free *MT* analysis (11.11–12.50 bits/s for 1D touch-pointing tasks) [43]. Therefore, we conclude that both participant groups appropriately followed our instruction on trying to balance speed and accuracy.

9.3 Internal and External Validity of Prediction Parameters

Because the main scope of our study did not include testing the external validity of the prediction parameters in equations (α and σ_a), it is sufficient that the observed and predicted success rates *internally* matched the participant group. Yet, it is still worth discussing the external validity of the prediction parameters for better understanding of the dual Gaussian distribution hypothesis.

Bi and Zhai measured the parameters of α and σ_a in their earlier study [9]. Those parameters were then used in Equations 8 and 10 to predict the success rates [10]. Because the participants in those two studies differed, the parameters of α and σ_a could have had *external* validity. This validated Bi et al.’s assumption: “Assuming finger size and shape do not vary drastically across users, σ_a could be used across users as an approximation” [8].

The top-left panel of Figure 12 shows the predicted success rates in the 1D tasks. In addition to the prediction data reported in Figures 3 and 5, we also computed the predicted success rates by using the σ_{obs_y} values measured in the 2D tasks of Experiments 3 and 4. The actual success rate in Experiment 1 under the condition of $W = 2$ mm was 71.55% (top-right panel of Figure 12), and those in Experiment 2 ranged from 71.73 to 77.60%. Therefore, we conclude that using the σ_{obs_y} values measured in the 2D tasks would allow us to predict more accurate success rates for 1D tasks. Here, using Bi and Zhai’s generic σ_{obs_y} value [10] allows us to predict the success rate (60.66%), but this is not as close as ours to the actual data. Note that three students participated on both days in our study; this is not a complete comparison as an external validity check.

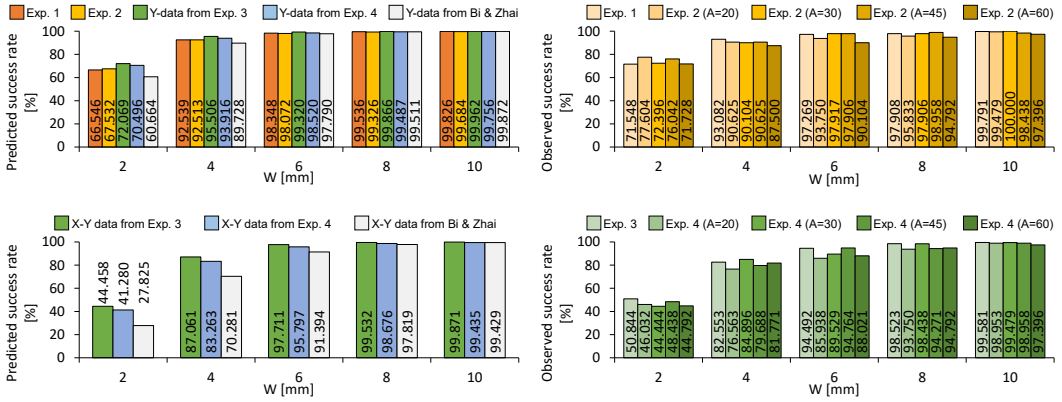


Fig. 12. Comparison of predicted (left) vs. observed (right) success rates from our data and Bi and Zhai's [10] for 1D (top) and 2D (bottom) tasks.

We also tried to determine whether the success rates in the 2D tasks could be predicted from Bi and Zhai's data, as shown in the bottom-left panel of Figure 12. Because Bi and Zhai's data for σ_a was larger than ours, their predicted success rates tended to be lower. Furthermore, because the actual success rate was over 50% for $W = 2$ mm in Experiment 3 (bottom-right panel of Figure 12), Bi and Zhai's prediction parameters could not be used to predict the success rates in our experiments. Note that using Bi and Zhai's prediction parameters for the index finger [10] instead of generic parameters would not influence this conclusion.

Possible explanations for why Bi and Zhai's parameters did not fit for our data include differences in user groups and devices. This result supports Bi, Li, and Zhai's hypothesis that σ_a may vary with an individual's finger size or motor impairment (e.g., tremor, or lack of) [8]. The fact that the model parameters α and σ_a can change depending on the user group and thus affect the success-rate prediction accuracy is an empirically demonstrated limitation on the generalizability of the dual Gaussian distribution hypothesis. This is one of the novel findings of our study, as it has never been shown with such evidence.

In the case that the properties of the main users of an app or the visitors to a website are already known (e.g., an education app that is mainly used by children 10–12 years old), this result suggests that designers should choose the appropriate participants for measuring the prediction parameters α and σ_a to further increase the prediction accuracy of the success rate. For example, Leitão and Silva listed various apps having large buttons and swipe widgets suitable for older adults (Figures 1–11 in [33]). In addition, if designers can conduct several experiments with various user groups, designing UIs differently according to the users would also be beneficial and has been already adopted. For example, on *YouTube Kids* [29], the button size is auto-personalized depending on the age listed in the user's account information. Our results can help with such optimization and personalization according to the characteristics of the target users.

9.4 Potential Improvements by Integrating the Target Distance

Because we observed that A significantly affected σ_{obsy} values in Experiments 2 and 4, the accuracy in predicting the success rate would be potentially improved by integrating the A factor. However, the results showed no strong benefits of using A for prediction, as summarized in Table 1.

Because the candidate formulations for modeling σ_{obs} values had different numbers of free parameters, we used the adjusted R^2 . We also compared models with the Akaike information

criterion (*AIC*) values [1]. This balanced the model complexity and the fitness to determine the comparatively best model. A model with a lower *AIC* value is a better one, and a model with an *AIC* that is greater than $AIC_{\text{minimum}} + 10$ is a significantly worse one. However, the *AIC* value tends to be smaller for a model with more free parameters if the number of measured data points is large (here, $N = 20 = 4_W \times 5_A$). To address this, we also used the Bayesian information criterion *BIC* [28]. A model with a lower *BIC* is better. The strengths for judging a significant difference in *BIC* are: 0–2 indicates no significant difference, 2–6 is positive, 6–10 is strong, and >10 is very strong [28].

9.4.1 Experiment 2. The results showed that *A* had significant main effects on σ_{obs_y} and the success rate. The interactions of $A \times W$ on σ_{obs_y} and the success rate were not significant; thus, *A* and *W* independently affected σ_{obs_y} . First, in Section 6.2, we reported the linear regression result using Equation 4 ($\sigma_{\text{obs}}^2 = \alpha W^2 + \sigma_a^2$) for $N = 20$ data points. This result is shown as Model #1 in Table 1. Second, to take the effect of *A* into account, we integrated Equation 4 ($\sigma_{\text{obs}}^2 = \alpha W^2 + \sigma_a^2$) and Equation 11 ($\sigma_{\text{obs}}^2 = \beta A^2 + \gamma$), and thus, we obtained:

$$\sigma_{\text{obs}}^2 = \alpha W^2 + \beta A^2 + \text{intercept} \quad (12)$$

where the intercept is $\sigma_a^2 + \gamma$. This addition is achieved because the variances of two independent normal distributions can be additive and the covariance of W^2 and A^2 is zero. The result is shown as Model #2 in Table 1. Compared with Model #1, the adjusted R^2 value increased from 0.8038 to 0.8125. However, according to the *AIC* and *BIC* criteria, there were no significant differences in predicting the σ_{obs_y} value. Regarding the success-rate prediction, the results were slightly better for Model #2 for all three metrics. Therefore, we found that Model #2 had a slightly better prediction accuracy over Model #1.

9.4.2 Experiment 4. The results showed that *A* had a significant main effect on σ_{obs_y} , but not for σ_{obs_x} and the success rate. Also, interactions of $A \times W$ were not found for σ_{obs_x} , σ_{obs_y} , and the success rate. In the same manner as with Experiment 2, we report the model fitness in terms of predicting σ_{obs_x} and σ_{obs_y} using Equations 4 and 12, as listed in Table 1.

In Section 8.1, we reported the results of using Equation 4, which ignores the effects of *A* on σ_{obs_x} and σ_{obs_y} . These results correspond to Models #3 and #4. Next, we took the effects of *A* into account, as shown as Models #5 and #6. Comparing these results, we found that using Equation 12 is effective for predicting σ_{obs_y} , but using Equation 4 is better for σ_{obs_x} . This would be because the *A* had a significant effect only for σ_{obs_y} . Thus, using Equation 12 for σ_{obs_y} while using Equation 4 for σ_{obs_x} is a more empirically sound approach. This result is shown in the bottom row in Table 1. However, the accuracy in predicting the success rate is worse than always using Equation 4.

In summary, an improvement in success-rate prediction was observed only for the results from Experiment 2. Still, for the results from Experiment 2, the improvement is not remarkable; the difference in the R^2 of the observed vs. predicted values was 0.01, the *MAE* difference was 0.03 points, and the maximum prediction error (diff_{max}) difference was 0.95 points (1.4%). Also, this improvement should come from the higher model fitness for σ_{obs_y} , but this is not evident according to the *AIC* and *BIC* criteria. These results prevent us from claiming that we should take the effect of *A* into account to predict the success rate. As we showed, integrating *A* to regress the σ_{obs} values did not improve the fitness significantly, and thus, we identified that Bi and Zhai’s decision not to test *A* as an independent variable in their experiment was a reasonable choice (e.g., for increasing the numbers of repetitions for each *W* condition).

Table 1. Model fitting and prediction accuracy using different formulations to predict the σ_{obs}^2 for the data from Experiments 2 and 4. diff_{max} is the maximum prediction error between the predicted versus observed success rates for $N = 20$ data points.

Exp.	Eq.	Variance (σ^2) fitting by linear regression			Success-rate prediction				
		Model	adj. R^2	AIC	BIC	R^2	MAE (%)	diff_{max} (points & %)	
2	4	(#1) $\sigma_{\text{obs}_y}^2 = 0.01915W^2 + 0.9543$	0.8038	16.73	13.73	0.9242	3.266	10.07 (14.9%)	
	12	(#2) $\sigma_{\text{obs}_y}^2 = 0.01915W^2 + 0.00008031A^2 + 0.8153$	0.8125	16.69	14.68	0.9349	3.236	9.057 (13.5%)	
3	4	(#3) $\sigma_{\text{obs}_x}^2 = 0.01105W^2 + 0.9227$	0.8102	-6.038	-9.043	0.9753	3.671	9.859 (10.30%)	
		(#4) $\sigma_{\text{obs}_y}^2 = 0.01885W^2 + 0.8366$	0.7200	25.28	22.27				
	12	(#5) $\sigma_{\text{obs}_x}^2 = 0.01105W^2 + 0.00001644A^2 + 0.8943$	0.8015	-4.295	-6.303	0.9620	3.744	11.02 (11.36%)	
		(#6) $\sigma_{\text{obs}_y}^2 = 0.01885W^2 + 0.0001776A^2 + 0.5291$	0.7955	19.85	17.84				
	A's effect only for $\sigma_{\text{obs}_x}^2$; use Model #3 for $\sigma_{\text{obs}_x}^2$ and Model #6 for $\sigma_{\text{obs}_y}^2$						0.9639	3.678	10.93 (11.28%)

9.5 Limitations and Future Work

Our findings are somewhat limited by the experimental conditions, such as the A and W values used in the tasks. In particular, much longer A values have been tested in touch-pointing studies, e.g., 20 cm [34]. Hence, our conclusions are limited to small screens. The limited range of A values provides one possible reason why we observed only two pairs having significant differences in σ_{obs} (between $A = 45$ and 60 mm in Experiment 2 and $A = 30$ and 60 mm in Experiment 4). In addition, Bi and Zhai measured prediction parameters for using both the thumb in a one-handed posture and the index finger [9], and they also measured the success rates in 1D pointing with a vertical bar target [10]. If we conduct user studies under such conditions, they would provide additional contributions in the future.

Our experiments required the participants to balance speed and accuracy. In other words, the participants could take their time if necessary. The success rate has been shown to vary nonlinearly depending on whether users try to shorten the operation time or aim carefully [52, 54, 55]. Our experimental instructions covered just one case among the various situations of touch selection.

10 CONCLUSION

We discussed the applicability of Bi and Zhai's success-rate prediction model [10] to pointing tasks starting on-screen. The potential concern about an on-screen start in such tasks was that the movement distance A is both implicitly and explicitly defined, and previous studies suggested that the A value would influence the endpoint variability. We empirically showed the validity of the model in four experiments. The prediction error was at most 10.07 points (14.9%) among 50 conditions in total. Also, we found that the effectiveness of integrating A to predict the success rates was limited. Our results indicate that designers and researchers can accurately predict the success rate by using a single model, regardless of whether a user taps a certain GUI item by moving a finger to the screen or keeping it close to the surface as in keyboard typing. Our findings will be beneficial for designing better touch GUIs and for automatically generating and optimizing UIs.

REFERENCES

- [1] Hirotugu Akaike. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Control* 19, 6 (Dec 1974), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- [2] Oscar Kin-Chung Au, Xiaojun Su, and Rynson W.H. Lau. 2014. LinearDragger: A Linear Selector for One-finger Target Acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (CHI '14). ACM, New York, NY, USA, 2607–2616. <https://doi.org/10.1145/2556288.2557096>

- [3] Daniel Avrahami. 2015. The Effect of Edge Targets on Touch Performance. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) (*CHI '15*). ACM, New York, NY, USA, 1837–1846. <https://doi.org/10.1145/2702123.2702439>
- [4] Shiri Azenkot and Shumin Zhai. 2012. Touch Behavior with Different Postures on Soft Smartphone Keyboards. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services* (San Francisco, California, USA) (*MobileHCI '12*). ACM, New York, NY, USA, 251–260. <https://doi.org/10.1145/2371574.2371612>
- [5] Gilles Bailly, Antti Oulasvirta, Timo Kötzing, and Sabrina Hoppe. 2013. MenuOptimizer: Interactive Optimization of Menu Systems. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) (*UIST '13*). ACM, New York, NY, USA, 331–342. <https://doi.org/10.1145/2501988.2502024>
- [6] W. D. A. Beggs, Jacqueline A. Andrew, Martha L. Baker, S. R. Dove, Irene Fairclough, and C. I. Howarth. 1972. The accuracy of non-visual aiming. *Quarterly Journal of Experimental Psychology* 24, 4 (1972), 515–523. <https://doi.org/10.1080/14640747208400311>
- [7] W. D. A. BEGGS, RUTH SAKSTEIN, and C. I. HOWARTH. 1974. The Generality of a Theory of the Intermittent Control of Accurate Movements. *Ergonomics* 17, 6 (1974), 757–768. <https://doi.org/10.1080/00140137408931422>
- [8] Xiaojun Bi, Yang Li, and Shumin Zhai. 2013. FFitts Law: Modeling Finger Touch with Fitts' Law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) (*CHI '13*). ACM, New York, NY, USA, 1363–1372. <https://doi.org/10.1145/2470654.2466180>
- [9] Xiaojun Bi and Shumin Zhai. 2013. Bayesian touch: a statistical criterion of target selection with finger touch. In *Proceedings of the ACM Symposium on User Interface Software and Technology* (*UIST '13*). 51–60. <https://doi.org/10.1145/2501988.2502058>
- [10] Xiaojun Bi and Shumin Zhai. 2016. Predicting Finger-Touch Accuracy Based on the Dual Gaussian Distribution Model. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) (*UIST '16*). ACM, New York, NY, USA, 313–319. <https://doi.org/10.1145/2984511.2984546>
- [11] Andy Cockburn, David Ahlström, and Carl Gutwin. 2012. Understanding performance in touch selections: Tap, drag and radial pointing drag with finger, stylus and mouse. *International Journal of Human-Computer Studies* 70, 3 (2012), 218 – 233. <https://doi.org/10.1016/j.ijhcs.2011.11.002>
- [12] Edward R.F.W. Crossman. 1956. *The speed and accuracy of simple hand movements*. Ph.D. Dissertation. University of Birmingham.
- [13] Peter Dixon. 2008. Models of accuracy in repeated-measures designs. *Journal of Memory and Language* 59, 4 (2008), 447–456.
- [14] Jan Eggers, Dominique Feillet, Steffen Kehl, Marc Oliver Wagner, and Bernard Yannou. 2003. Optimization of the keyboard arrangement problem using an Ant Colony algorithm. *European Journal of Operational Research* 148, 3 (2003), 672–686. [https://doi.org/10.1016/S0377-2217\(02\)00489-7](https://doi.org/10.1016/S0377-2217(02)00489-7)
- [15] Paul M. Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47, 6 (1954), 381–391. <https://doi.org/10.1037/h0055392>
- [16] Krzysztof Gajos and Daniel S. Weld. 2004. SUPPLE: Automatically Generating User Interfaces. In *Proceedings of the 9th International Conference on Intelligent User Interfaces* (Funchal, Madeira, Portugal) (*IUI '04*). ACM, New York, NY, USA, 93–100. <https://doi.org/10.1145/964442.964461>
- [17] Khai-Chung Gan and Errol R. Hoffmann. 1988. Geometrical conditions for ballistic and visually controlled movements. *Ergonomics* 31, 5 (1988), 829–839. <https://doi.org/10.1080/00140138808966724>
- [18] Julien Gori, Olivier Rioul, and Yves Guiard. 2018. Speed-Accuracy Tradeoff: A Formal Information-Theoretic Transmission Scheme (FITTS). *ACM Trans. Comput.-Hum. Interact.* 25, 5, Article 27 (Sept. 2018), 33 pages. <https://doi.org/10.1145/3231595>
- [19] Niels Henze, Enrico Rukzio, and Susanne Boll. 2012. Observational and Experimental Investigation of Typing Behaviour Using Virtual Keyboards for Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) (*CHI '12*). ACM, New York, NY, USA, 2659–2668. <https://doi.org/10.1145/2207676.2208658>
- [20] Errol R. Hoffmann. 2016. Critical Index of Difficulty for Different Body Motions: A Review. *Journal of Motor Behavior* 48, 3 (2016), 277–288. <https://doi.org/10.1080/00222895.2015.1090389>
- [21] Christian Holz and Patrick Baudisch. 2010. The Generalized Perceived Input Point Model and How to Double Touch Accuracy by Extracting Fingerprints. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) (*CHI '10*). ACM, New York, NY, USA, 581–590. <https://doi.org/10.1145/1753326.1753413>
- [22] Christian Holz and Patrick Baudisch. 2011. Understanding Touch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (*CHI '11*). ACM, New York, NY, USA, 2501–2510. <https://doi.org/10.1145/1978942.1979308>
- [23] Jin Huang and Byungjoo Lee. 2019. Modeling Error Rates in Spatiotemporal Moving Target Selection. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (*CHI EA '19*).

- Association for Computing Machinery, New York, NY, USA, Article LBW2411, 6 pages. <https://doi.org/10.1145/3290607.3313077>
- [24] Jin Huang, Feng Tian, Xiangmin Fan, Xiaolong (Luke) Zhang, and Shumin Zhai. 2018. Understanding the Uncertainty in 1D Unidirectional Moving Target Selection. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI '18*). Association for Computing Machinery, New York, NY, USA, Article 237, 12 pages. <https://doi.org/10.1145/3173574.3173811>
- [25] Jin Huang, Feng Tian, Xiangmin Fan, Xiaolong (Luke) Zhang, and Shumin Zhai. 2018. Understanding the Uncertainty in 1D Unidirectional Moving Target Selection. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI '18*). ACM, New York, NY, USA, Article 237, 12 pages. <https://doi.org/10.1145/3173574.3173811>
- [26] Jin Huang, Feng Tian, Nianlong Li, and Xiangmin Fan. 2019. Modeling the Uncertainty in 2D Moving Target Selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) (*UIST '19*). Association for Computing Machinery, New York, NY, USA, 1031–1043. <https://doi.org/10.1145/3332165.3347880>
- [27] Human Factors and Ergonomics Society 2007. American National Standard for Human Factors Engineering of Computer Workstations (ANSI/HFES Standard No. 100-2007).
- [28] Robert E. Kass and Adrian E. Raftery. 1995. Bayes Factors. *J. Amer. Statist. Assoc.* 90, 430 (1995), 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- [29] YouTube Kids. 2015. Retrieved August 28, 2020 from <https://apps.apple.com/us/app/youtube-kids/id936971630>, or <https://play.google.com/store/apps/details?id=com.google.android.apps.youtube.kids>.
- [30] Byungjoo Lee, Sunjun Kim, Antti Oulasvirta, Jong-In Lee, and Eunji Park. 2018. Moving Target Selection: A Cue Integration Model. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (*CHI '18*). ACM, New York, NY, USA, Article 230, 12 pages. <https://doi.org/10.1145/3173574.3173804>
- [31] Byungjoo Lee and Antti Oulasvirta. 2016. Modelling Error Rates in Temporal Pointing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '16*). ACM, New York, NY, USA, 1857–1868. <https://doi.org/10.1145/2858036.2858143>
- [32] Injung Lee, Sunjun Kim, and Byungjoo Lee. 2019. Geometrically Compensating Effect of End-to-End Latency in Moving-Target Selection Games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (*CHI '19*). ACM, New York, NY, USA. <https://doi.org/10.1145/3290605.3300790>
- [33] Roxanne Leitão and Paula Alexandra Silva. 2012. Target and Spacing Sizes for Smartphone User Interfaces for Older Adults: Design Patterns Based on an Evaluation with Users. In *Proceedings of the 19th Conference on Pattern Languages of Programs* (Tucson, Arizona) (*PLoP '12*). The Hillside Group, USA, Article 5, 13 pages. <http://dl.acm.org/citation.cfm?id=2821679.2831275>
- [34] Yuexing Luo and Daniel Vogel. 2014. Crossing-based Selection with Direct Touch Input. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (*CHI '14*). ACM, New York, NY, USA, 2627–2636. <https://doi.org/10.1145/2556288.2557397>
- [35] I. Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction* 7, 1 (1992), 91–139. https://doi.org/10.1207/s15327051hci0701_3
- [36] Blanca Mena, M José, Rafael Alarcón, Jaume Arnau Gras, Roser Bono Cabré, and Rebecca Bendayan. 2017. Non-normal data: Is ANOVA still a valid option? *Psicothema*, 2017, vol. 29, num. 4, p. 552–557 (2017).
- [37] David E. Meyer, Richard A. Abrams, Sylvan Kornblum, Charles E. Wright, and J. E. Keith Smith. 1988. Optimality in human motor performance: ideal control of rapid aimed movements. *Psychological Review* 95, 3 (1988), 340–370. <https://doi.org/10.1037/0033-295X.95.3.340>
- [38] David E. Meyer, J. E. Keith Smith, and Charles E. Wright. 1982. Models for the speed and accuracy of aimed movements. *Psychological Review* 89, 5 (1982), 449–482. <https://doi.org/10.1037/0033-295X.89.5.449>
- [39] Tomer Moscovich. 2009. Contact Area Interaction with Sliding Widgets. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology* (Victoria, BC, Canada) (*UIST '09*). ACM, New York, NY, USA, 13–22. <https://doi.org/10.1145/1622176.1622181>
- [40] Jeffrey Nichols, Brad A. Myers, and Kevin Litwack. 2004. Improving Automatic Interface Generation with Smart Templates. In *Proceedings of the 9th International Conference on Intelligent User Interfaces* (Funchal, Madeira, Portugal) (*IUI '04*). ACM, New York, NY, USA, 286–288. <https://doi.org/10.1145/964442.964507>
- [41] Jakob Nielsen. 2010. Scrolling and Attention (Original Research Study). Retrieved August 28, 2020 from <https://www.nngroup.com/articles/scrolling-and-attention-original-research/>.
- [42] Eunji Park and Byungjoo Lee. 2018. Predicting Error Rates in Pointing Regardless of Target Motion. arXiv:arXiv:1806.02973 <https://arxiv.org/abs/1806.02973>
- [43] Esben Warming Pedersen and Kasper Hornbæk. 2012. An Experimental Comparison of Touch Interaction on Vertical and Horizontal Surfaces. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (Copenhagen, Denmark) (*NordiCHI '12*). ACM, New York, NY, USA, 370–379. <https://doi.org/10.1145/>

2399016.2399074

- [44] Katrin Plaumann, Milos Babic, Tobias Drey, Witali Hepting, Daniel Stooss, and Enrico Rukzio. 2018. Improving Input Accuracy on Smartphones for Persons Who Are Affected by Tremor Using Motion Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 156 (Jan. 2018), 30 pages. <https://doi.org/10.1145/3161169>
- [45] R. L. Potter, L. J. Weldon, and B. Shneiderman. 1988. Improving the Accuracy of Touch Screens: An Experimental Evaluation of Three Strategies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Washington, D.C., USA) (*CHI '88*). ACM, New York, NY, USA, 27–32. <https://doi.org/10.1145/57167.57171>
- [46] Richard A. Schmidt, Howard N. Zelaznik, Bob Hawkins, James S. Frank, and J. T. Quinn. 1979. Motor-output variability: a theory for the accuracy of rapid motor acts. *Psychological review* 86, 5 (1979), 415–451.
- [47] R. William Soukoreff and I. Scott MacKenzie. 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies* 61, 6 (2004), 751–789. <https://doi.org/10.1016/j.ijhcs.2004.09.001>
- [48] Daniel Vogel and Patrick Baudisch. 2007. Shift: A Technique for Operating Pen-based Interfaces Using Touch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '07*). ACM, New York, NY, USA, 657–666. <https://doi.org/10.1145/1240624.1240727>
- [49] Stephen A. Wallace and Karl M. Newell. 1983. Visual control of discrete aiming movements. *The Quarterly Journal of Experimental Psychology Section A* 35, 2 (1983), 311–321. <https://doi.org/10.1080/14640748308402136>
- [50] Feng Wang and Xiangshi Ren. 2009. Empirical Evaluation for Finger Input Properties in Multi-touch Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) (*CHI '09*). ACM, New York, NY, USA, 1063–1072. <https://doi.org/10.1145/1518701.1518864>
- [51] Daryl Weir, Simon Rogers, Roderick Murray-Smith, and Markus Löchtfeld. 2012. A User-specific Machine Learning Approach for Improving Touch Accuracy on Mobile Devices. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology* (Cambridge, Massachusetts, USA) (*UIST '12*). ACM, New York, NY, USA, 465–476. <https://doi.org/10.1145/2380116.2380175>
- [52] Wayne A. Wickelgren. 1977. Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica* 41, 1 (1977), 67–85. [https://doi.org/10.1016/0001-6918\(77\)90012-9](https://doi.org/10.1016/0001-6918(77)90012-9)
- [53] Daniel Wigdor, Sarah Williams, Michael Cronin, Robert Levy, Katie White, Maxim Mazeev, and Hrvoje Benko. 2009. Ripples: Utilizing Per-contact Visualizations to Improve User Interaction with Touch Displays. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology* (Victoria, BC, Canada) (*UIST '09*). ACM, New York, NY, USA, 3–12. <https://doi.org/10.1145/1622176.1622180>
- [54] Jacob O. Wobbrock, Edward Cutrell, Susumu Harada, and I. Scott MacKenzie. 2008. An Error Model for Pointing Based on Fitts' Law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) (*CHI '08*). ACM, New York, NY, USA, 1613–1622. <https://doi.org/10.1145/1357054.1357306>
- [55] Jacob O. Wobbrock, Alex Jansen, and Kristen Shinohara. 2011. Modeling and Predicting Pointing Errors in Two Dimensions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (*CHI '11*). ACM, New York, NY, USA, 1653–1656. <https://doi.org/10.1145/1978942.1979183>
- [56] Jacob O. Wobbrock, Kristen Shinohara, and Alex Jansen. 2011. The Effects of Task Dimensionality, Endpoint Deviation, Throughput Calculation, and Experiment Design on Pointing Measures and Models. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) (*CHI '11*). ACM, New York, NY, USA, 1639–1648. <https://doi.org/10.1145/1978942.1979181>
- [57] Shota Yamanaka. 2018. Effect of Gaps with Penal Distractors Imposing Time Penalty in Touch-pointing Tasks. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Barcelona, Spain) (*MobileHCI '18*). ACM, New York, NY, USA, 8. <https://doi.org/10.1145/3229434.3229435>
- [58] Shota Yamanaka. 2018. Risk Effects of Surrounding Distractors Imposing Time Penalty in Touch-Pointing Tasks. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces* (Tokyo, Japan) (*ISS '18*). ACM, New York, NY, USA, 129–135. <https://doi.org/10.1145/3279778.3279781>
- [59] Koji Yatani, Kurt Partridge, Marshall Bern, and Mark W. Newman. 2008. Escape: A Target Selection Technique Using Visually-cued Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) (*CHI '08*). ACM, New York, NY, USA, 285–294. <https://doi.org/10.1145/1357054.1357104>
- [60] Chun Yu, Hongyi Wen, Wei Xiong, Xiaojun Bi, and Yuanchun Shi. 2016. Investigating Effects of Post-Selection Feedback for Acquiring Ultra-Small Targets on Touchscreen. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (*CHI '16*). ACM, New York, NY, USA, 4699–4710. <https://doi.org/10.1145/2858036.2858593>
- [61] Howard N. Zelaznik, Susan Mone, George P. McCabe, and Christopher Thaman. 1988. Role of temporal and spatial precision in determining the nature of the speed-accuracy trade-off in aimed-hand movements. *Journal of Experimental Psychology: Human Perception and Performance* 14, 2 (1988), 221–230. <https://doi.org/10.1037/0096-1523.14.2.221>
- [62] Shumin Zhai. 2020. Personal communication, September 4th.

- [63] Shumin Zhai, Jing Kong, and Xiangshi Ren. 2004. Speed-accuracy tradeoff in Fitts' law tasks: on the equivalency of actual and nominal pointing precision. *International Journal of Human-Computer Studies* 61, 6 (2004), 823–856. <https://doi.org/10.1016/j.ijhcs.2004.09.007>

Received June 2020; revised August 2020; accepted September 2020