Verifying Finger-Fitts Models for Normalizing Subjective Speed-Accuracy Biases

SHOTA YAMANAKA, LY Corporation, Japan
HIROKI USUBA, LY Corporation, Japan
YOSUKE OBA, Meiji University, Japan
TAIKI KINOSHITA, Meiji University, Japan
RYUTO TOMIHARI, Meiji University, Japan
NOBUHITO KASAHARA, Meiji University, Japan
HOMEI MIYASHITA, Meiji University, Japan

Previous studies on the Finger-Fitts law (FFitts law) are lacking in sufficient experiments to verify its inherent potential. Since the FFitts law is originally a modified version of the *effective width method* to normalize speed-accuracy biases, the model fit would improve if multiple biases were mixed together and the throughputs would be more stable than using the nominal target width. In this study, we conduct an experiment in which participants tap 1D-bar and 2D-circular targets under three subjective biases: balancing the speed and accuracy, emphasizing speed, and emphasizing accuracy when they perform the tasks. The results showed that applying the effective width to Ko et al.'s refined FFitts law, which represents the touch ambiguity with a free parameter, was the most successful in normalizing biases. Reanalyzing another dataset on ray-casting pointing also led to the same conclusion. We thus recommend using Ko et al.'s model with effective width when researchers compare several experimental conditions such as devices and user groups.

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

Additional Key Words and Phrases: finger-touch pointing, time prediction model, effective width, FFitts law

ACM Reference Format:
Shota Yamanaka, Hiroki Usuba, Yosuke Oba, Taiki Kinoshita, Ryuto Tomihari, Nobuhito Kasahara, and Homei Miyashita. 2024. Verifying Finger-Fitts Models for Normalizing Subjective Speed-Accuracy Biases. *Proc. ACM Hum.-Comput. Interact.* 8, MHCI, Article 285 (September 2024), 24 pages. https://doi.org/10.1145/3676532

1 Introduction

1.1 Background

Pointing at targets is one of the most frequent operations in PC or smartphone usage. In the HCI field, deriving user-performance models to predict task outcomes is a core topic, and Fitts' law is widely utilized to predict the movement time *MT* in pointing [19]. In addition, Fitts' law is used to calculate a usability metric called throughput *TP*, which is used to compare the performance of different experimental factors such as devices and user groups [47].

Authors' Contact Information: Shota Yamanaka, syamanak@lycorp.co.jp, LY Corporation, Tokyo, Japan; Hiroki Usuba, hrky0206@gmail.com, LY Corporation, Tokyo, Japan; Yosuke Oba, bonscow@gmail.com, Meiji University, Tokyo, Japan; Taiki Kinoshita, kinop.tofu@gmail.com, Meiji University, Tokyo, Japan; Ryuto Tomihari, ryuutomi1110@gmail.com, Meiji University, Tokyo, Japan; Nobuhito Kasahara, k.nobu00714@outlook.jp, Meiji University, Tokyo, Japan; Homei Miyashita, homei@homei.com, Meiji University, Tokyo, Japan.

ACM 2573-0142/2024/9-ART285
https://doi.org/10.1145/3676532

Fitts' law holds for various devices, including touchscreens [10, 61] and hand-held controllers for virtual reality (VR) environments or wall-sized displays [6, 31]. However, when users operate touchscreens, several issues may prevent them from accurately tapping the intended target, such as occlusion of the target or deformation of the finger pad [27, 28]. Since the target size (or width W) cannot be fully utilized in such situations, Bi et al. proposed the Finger-Fitts law (FFitts law), which modifies Fitts' law to integrate the ambiguity of finger-touch coordinates [10].

Although several replications on FFitts law have been conducted, they are lacking in sufficient experiments to verify the inherent potential of this model. FFitts law is modified from the effective width method [47, 53], which adjusts the target size by the endpoint variability *a posteriori*. The benefit of using the effective width W_e is that it normalizes the subjective biases when researchers analyze the data exhibited by participants operating with multiple speed-accuracy priorities. For example, Zhai [63] and Yamanaka [59] asked the participants to adhere to three speed-accuracy biases, and we will employ similar biases in our experiment as follows:

- *Neutral*: Perform the task as quickly and accurately as possible.
- *Accurate*: Perform the task so as not to make errors as much as possible without worrying about the duration.
- *Fast*: Perform the task as quickly as possible without worrying about making errors.

However, in previous studies on FFitts law, experiments were conducted with only one bias [10, 32, 35, 36, 50, 53, 55, 61].[1]

1.2 Research Hypothesis and Contribution

There are two metrics to identify the benefit of using W_e instead of the nominal W [42, 62, 63]: (a) when analyzing data from operations with multiple biases mixed together, the model fit is better, and (b) *TP*s are more stable across different biases, i.e., the *TP*s under different speed-accuracy bias conditions fall within a narrower range. Because FFitts law originally improved on the effective width method, our research hypothesis for touch pointing is that **"Combining FFitts law with W_e should further improve the normalization capability in terms of (a) and (b) above, rather than using either FFitts law or W_e alone."**

If using only W_e without FFitts law can normalize the impact of bias, researchers can use the simple model formulation to calculate *TP*, which would support previous studies that compared user performance using only W_e. Alternatively, it might be sufficient to use only FFitts law without W_e, or to use a combination of FFitts law and W_e. On the basis of the results of previous studies, we can probably assume that the best choice is to use both FFitts law and W_e, since using W_e can normalize the effects of bias and FFitts law should fit the *MT* data in touch operations.

However, there is no evidence to indicate which of these potential combinations is the best. It is scientifically inappropriate for researchers to make decisions about whether to combine FFitts law and W_e purely on speculation without empirical data. Until now, there has been neither evidence for doing so nor counter-examples, e.g., researchers might have reached a conclusion that using only FFitts law is sufficient to achieve (a) and (b) above, and combining FFitts law and W_e gave no significant improvement. This lack of evidence is what motivated us to conduct the current work.

In this study, we conducted two touch-pointing tasks and reanalyzed existing datasets on pointing in VR [8]. We will justify the application of FFitts law to pointing in VR later. Our findings led to the same conclusion for two touch-pointing tasks and the ray-casting method, namely, that we should combine FFitts law and W_e for normalizing speed-accuracy biases.

[1] We surveyed all articles that cited the original FFitts law paper [10] in the ACM Digital Library and Google Scholar as of February 15th, 2024.

Our contribution is thus to experimentally demonstrate that using both FFitts law and W_e can better normalize the speed-accuracy biases. While researchers have traditionally used W_e alone for comparative experiments with touchscreens and ray casting [5, 18, 21], this work will enable researchers to more fairly compare the performances of different factors (e.g., devices, interaction techniques, and user groups).

2 Related Work

2.1 Movement Time Models

Fitts' law predicts the MT based on the target distance A and its width W with regression constants (a and b) [19, 47]:

$$\text{Baseline Fitts' law: } MT = a + b \cdot ID \quad \text{and} \quad ID = \log_2\left(\frac{A}{W} + 1\right), \tag{1}$$

where ID is the index of difficulty. In a typical user experiment, participants are asked to "perform the task as quickly and as accurately as possible" to balance the speed and accuracy, which we call the *Neutral* bias instruction.

Target pointing has been utilized in various comparative experiments in the HCI field. Examples include comparing the performance of devices [12, 20], user attributes [18, 45], and selection techniques [14, 23]. In such experiments, even when researchers give participants the *Neutral* instruction, bias may occur towards speed or accuracy. For example, in a device comparison experiment where a mouse has MT = 1000 ms and the error rate ER = 7% while a trackball has MT = 1300 ms and ER = 3%, it is difficult to determine which has the higher performance. If both devices exhibit the same ER, we can conclude that the device with the smaller MT has a higher performance, but this is quite a rare case.

The effective width method enables such an equivalent ER, as the W_e is adjusted so that ER becomes 4% [47]. The W_e is calculated by the variance σ^2 of the endpoint (e.g., click- or tap-point coordinate) distribution:

$$W_e = \sqrt{2\pi e \sigma^2} = 4.133\sigma. \tag{2}$$

We can then compare TPs for both devices [47]:

$$TP = \overline{ID/MT}. \tag{3}$$

TP indicates the difficulty level performed per unit of time. By applying W_e to ID in Eq. 3, we can fairly compare the performance of several devices having the adjusted speed-accuracy bias.[2]

2.2 Finger-Fitts Law Models

According to Bi et al.'s dual Gaussian distribution model, the touch-point variability σ consists of two components [10]. One is the relative component (σ_r) affected by the speed-accuracy tradeoff of the users, i.e., the faster the users conduct the task, the wider the distribution becomes. The other is the absolute component (σ_a), which represents the precision of the finger. Assuming that the two components are independent, the touch-point variability is calculated as follows:

$$\sigma^2 = \sigma_r^2 + \sigma_a^2. \tag{4}$$

Originally, σ used for W_e (Eq. 2) indicated the endpoint variability affected by the speed-accuracy tradeoff, which is represented by σ_r in Eq. 4. Thus, Fitts' law for finger-pointing (Bi et al.'s FFitts

[2]Another extensively used definition of TP is $1/b$. The bias-normalization results of this formulation are included in the supplementary materials.

law [10]) is

$$\text{Bi et al.'s FFitts law: } MT = a + b\log_2\left(\frac{A}{4.133\sqrt{\sigma^2 - \sigma_a^2}} + 1\right). \tag{5}$$

σ_a is measured by a *calibration task* in which users repeatedly tap an extremely small target, and the endpoint variability represents σ_a [10]. Alternatively, since the squares of σ and W are linearly related (σ^2 = constant $\times W^2 + \sigma_a^2$), σ_a can be obtained from the intercept of this regression [11].

To compute σ and σ_a, either univariate or bivariate standard deviations (SD_x and SD_{xy}, respectively) of tap coordinates can be used. For SD_x, only endpoints projected onto the task axis (i.e., the direction from the start to the target) are used. For SD_{xy}, endpoints on a 2D plane are used. Even for 2D Fitts' law tasks, using SD_x is allowed [47, 54].

In both calibration and intercept methods, σ^2 is sometimes smaller than σ_a^2 [61]. In this case, the radicand in Eq. 5 becomes negative, and FFitts law cannot hold. Ko et al.'s modified FFitts law resolved this issue [35]:

$$\text{Ko et al.'s FFitts law: } MT = a + b\log_2\left(\frac{A}{\sqrt{W^2 - c^2}} + 1\right), \tag{6}$$

where c is a free parameter. Ko et al. utilized W instead of W_e because "participants respect the spatial constraint set by the task parameters" [35]. Yamanaka and Usuba found that Ko et al.'s FFitts law without root shows a good fit [61]:

$$\text{No-root Ko et al.'s FFitts law: } MT = a + b\log_2\left(\frac{A}{W - c} + 1\right). \tag{7}$$

In actuality, Welford proposed almost the same formulations as Eqs. 6 and 7 [53], with the differences being to use W_e instead of W and to use "+0.5" instead of "+1". His purpose to derive these models was to absorb the effect of hand tremor when participants operated a stylus in pointing tasks. Hence, the concepts for FFitts law and Welford's models are the same in terms of adjusting the target size that cannot be fully utilized.

Experiments by Bi et al. [10], Ko et al. [36], Yamanaka et al. [61], and Welford [53] have all reported that using W instead of W_e results in a better model fit, but these all used a single type of instruction (*Neutral*). Considering the original derivation basis of FFitts law, which involved subtracting a small value from W_e in the effective width method to absorb finger-touch ambiguity, the inherent potential of FFitts law lies in normalizing the speed-accuracy bias. This can be verified in experiments with multiple biases to measure MT data, where (a) the model fit is high even when regressing MTs from all bias conditions mixed together, and (b) the TPs are stable across biases. However, researchers have not yet undertaken such experiments.

2.3 Subjective Speed-Accuracy Biases

2.3.1 Effective Width Improves Model Fit when Mixing MT Data from Multiple Biases. Zhai et al. compared three biases in stylus-based pointing: to tap the targets "as accurately as possible" (*Accurate*), "as accurately as possible and as fast as possible" (*Neutral*), and "as fast as possible" (*Fast*) [63]. They found that using W_e showed higher R^2 when data from all bias conditions were regressed together, while using W yielded a higher R^2 for a single bias. These results were consistently found in their two additional experiments using two and five biases [63] and Yamanaka et al.'s studies [59, 62].

Fitts' law and its variations were designed to predict MT data measured under a single fixed condition, e.g., one group of participants using one device following one type of instruction. Therefore, it is out of scope to fit MT data from multiple instructions' operations mixed together. Nevertheless, according to Zhai et al. [63], using W_e for fitting resulted in a lower ID in the

Fast condition (where larger endpoint variability increased W_e) and a higher *ID* in the *Accurate* condition (where smaller endpoint variability reduced W_e). Thus, while biases affected *MT*, using W_e appropriately adjusted the *ID* values, which stabilized the intercept a and slope b when fitting *MT* for each bias. This enabled for a higher R^2 when mixing *MT* data from multiple biases than using W.

Zhai et al. thus claimed to use R^2 in a multiple-bias *MT* regression to determine whether W or W_e more effectively smooths out the effects of biases, and we will use this methodology. However, models with more free parameters tend to fit better to widely scattered *MT* data points. Therefore, we will employ adjusted R^2 and *AIC* [2], as well as cross-validation, which were not used by Zhai et al., to draw conclusions more fairly.

2.3.2 Effective Width Yields More Stable Throughput. MacKenzie and Isokoski conducted a mouse-based experiment with three biases and found that using W_e made the *TP*s close to each other, ranging from 5.67 to 5.73 bits/s (i.e., a difference of less than 1%) [42]. In contrast, Olafsdottir et al. tested two additional conditions, *max speed* and *max accuracy* [44], and found that *TP* ranged from approximately 6 to 10 bits/s (42% difference), suggesting that using W_e for *TP* does not normalize the speed-accuracy bias. However, the bias we examine in this study is subjectively invoked when participants perform pointing tasks in (e.g.) a standard device- or participant-comparison experiment. Therefore, unnecessarily excessive bias instructions such as *max accuracy* (i.e., "Always click on the same pixel" [44]) are out-of-scope.

To evaluate the bias-normalization capability, Yamanaka et al. defined the *TP* difference as $100\% \times (TP_{max} - TP_{min})/TP_{max}$, where TP_{max} and TP_{min} are the maximum and minimum *TP*s among the tested bias conditions [62]. If a model perfectly normalizes the biases, the *TP* difference is 0%. They showed that using W_e instead of W decreases the *TP* difference, which indicates a greater normalization capability of W_e.

3 Touch-Pointing Experiment

3.1 Tasks

We conducted four tasks: (1) 1D calibration task, (2) 1D Fitts task, (3) 2D calibration task, and (4) 2D Fitts task, all under the three bias conditions. The 1D calibration task was conducted for measuring σ_a in FFitts law for 1D. The participants repeatedly tapped a one-pixel horizontal-bar target that appeared in a random position. When the tapped position was within the target, a sound to indicate success was played; otherwise, a failure sound was played. The next target then appeared regardless of the success or failure of the trial.

In the 1D Fitts task, participants alternately tapped two horizontal-bar targets (Fig. 1a). The current target was orange, and the non-target was gray. The first target was the top one. The distance between the targets and their widths were A and W, respectively. If the tap position was outside the target, the color of the targets did not change, and participants re-aimed at the same target. Sound feedback was given depending on the success or failure for each tap.

The 2D calibration task was conducted for measuring σ_a in FFitts law for 2D. Participants aimed at the center of a 20-mm-wide crosshair target that appeared in a random position. Other than the shape of the target, this task was the same as the 1D calibration task. In both calibration tasks, the target was set to appear 20 mm from the inside edges of the screen, as targets located too close to the edge are known to worsen the touch performance [3].

In the 2D Fitts task, circular targets were arranged in a circle (cf. the ISO 9241-9 task [47]). Participants tapped the targets one after another in the order depicted in Fig. 1b. The color settings and sounds were the same as in the 1D task.

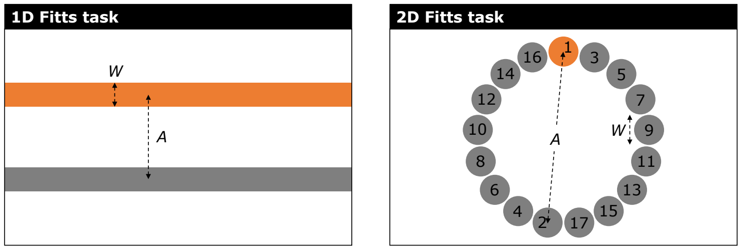


Fig. 1. 1D Fitts and 2D Fitts tasks.

3.2 Participants and Experimental System

Sixteen local university students participated in the experiment (age: *Mean* = 20.6, *SD* = 1.37). All participants were right-handed and were accustomed to touch interaction. They were asked to tap the targets with their right index finger. The experiment time was assumed to be 30 minutes, and the participants received 10.47 USD as a reward.

We used a desktop PC (Intel Core i9-12900KF, GeForce RTX 3070 Ti, 32 GB RAM, Windows 10 Home) and an external touch display (IK-KC209, 10-point multi touch, 13.3 inches, 1920 × 1080 pixels, 60 Hz). We set the display flat on a table and disabled Windows' touch feedback. The experimental system was developed in JavaScript.

3.3 Design and Procedure

In the Fitts tasks, we tested two As (26 and 60 mm) and three Ws (3, 5.5, and 8.5 mm). Fitts' ID ranged 2.02–4.39 bits. We chose a wider and higher ID range than several previous FFitts law studies (1.92–3.75 bits [10, 55]) to obtain more reliable model fits with regressions, while the range was narrower than another study (1.58–4.95 bits [61]) to shorten the experimental duration for examining three bias conditions. The order of the target conditions was random.

We tested three *Bias* conditions: *Neutral*, *Accurate*, and *Fast* (see Section 1.1). The instruction was always displayed at the top of the screen as a reminder.

Participants were divided into two groups: one conducted the 1D tasks first and the other did the 2D tasks first. Both groups performed the calibration tasks first and then the Fitts tasks. Regarding the order of *Bias*, all participants conducted *Neutral* first to help them conduct the task more accurately or faster in the second and third bias conditions (*Accurate* or *Fast*) [34, 60]. The order of *Accurate* and *Fast* was balanced.

In both calibration tasks, participants performed tapping 32 times for each *Bias*. Thus, the total number of trials was 1,536 ($3_{Bias} \times 32_{\text{repetitions}} \times 16_{\text{participants}}$). In both Fitts tasks, first, the participants conducted a *practice* block whose condition was fixed to A = 36 and W = 4.5 mm (ID = 3.20 bits) to set a medium difficulty level. In the practice and main blocks, participants repeated each condition 16 times. Thus, in the 1D task, they conducted eight round trips between the targets, and in the 2D task, they finished the trial on the 17th target. The total number of trials was 4,608 ($2_A \times 3_W \times 3_{Bias} \times 16_{\text{repetitions}} \times 16_{\text{participants}}$) in both Fitts tasks.

4 Results of Touch-Pointing Tasks

We removed tap positions that were 15 mm away from the target center as outliers to exclude accidental operations such as double-tapping the previous target, while we did not employ participant-level outliers. These criteria were consistent with previous studies [11, 61]. We removed seven outliers in the 1D calibration task (0.46%), ten in the 1D Fitts task (0.22%), eight in the 2D calibration task

Table 1. 1D: $\sigma^2 - \sigma_a^2$ and $\sigma^2 - intercept$ for each condition. σ_a is obtained from the calibration task. *intercept* is obtained from the regression intercept. The red values are negative, and the target size in FFitts law is undefined.

		Neutral $\sigma_a^2 = 1.01$, $intercept = 0.999$			*Accurate* $\sigma_a^2 = 0.814$, $intercept = 0.516$			*Fast* $\sigma_a^2 = 2.39$, $intercept = 2.21$		
A	W	σ^2	$\sigma^2 - \sigma_a^2$	$\sigma^2 - intercept$	σ^2	$\sigma^2 - \sigma_a^2$	$\sigma^2 - intercept$	σ^2	$\sigma^2 - \sigma_a^2$	$\sigma^2 - intercept$
26	3	0.977	−0.0334	0.00205	0.518	−0.297	−0.363	1.85	−0.544	−0.0220
26	5.5	1.16	0.150	0.00333	0.519	−0.296	0.225	2.44	0.0438	0.162
26	8.5	1.78	0.774	0.211	0.727	−0.0879	0.0189	2.23	−0.162	0.785
60	3	1.19	0.178	0.0329	0.549	−0.266	0.176	2.39	−0.00523	0.190
60	5.5	1.51	0.496	0.295	0.811	−0.00361	1.65	3.87	1.47	0.507
60	8.5	1.72	0.713	0.350	0.865	0.0507	2.35	4.56	2.17	0.724

(0.52%), and 22 in the 2D Fitts task (0.48%). To compute σ and σ_a, we used univariate SD_x for 1D and bivariate SD_{xy} for 2D tasks, as in previous FFitts law studies [10, 61].

4.1 1D Task

4.1.1 Movement Time, Error Rate, and σ_a. To check whether the participants followed each *Bias* instruction, we systematically ran repeated-measures ANOVA, as ANOVA is robust regardless of data distribution (e.g., violating the normality assumption) [17, 43]. For the Fitts task, dependent variables were *MT* (time from successfully tapping previous target to tapping the screen to aim for the current target including error trials) and *ER* (trials with one or more errors divided by the number of repetitions), and independent variables were A, W, and *Bias*. For the calibration task, the dependent variable was σ_a^2 and the independent variable was *Bias*. We used the Bonferroni p-value correction. If Mauchly's sphericity test was rejected, we used the Greenhouse–Geisser correction.

Regarding *MT*, we found the main effect for *Bias* ($F_{1.43,21.4} = 54.4$, $p < 0.001$, $\eta_p^2 = 0.784$). On average, we obtained 562 ms for *Neutral*, 777 ms for *Accurate*, and 397 ms for *Fast*, and all pairwise comparisons were significant ($p < 0.001$). Regarding *ER*, we found the main effect for *Bias* ($F_{2,30} = 39.4$, $p < 0.001$, $\eta_p^2 = 0.724$): 7.90% for *Neutral*, 2.80% for *Accurate*, and 15.9% for *Fast*, and all pairwise comparisons were significant ($p < 0.01$). These findings indicate that the participants accurately followed each *Bias*.

Regarding σ_a^2, we found the main effect for *Bias* ($F_{1.22,18.2} = 17.6$, $p < 0.001$, $\eta_p^2 = 0.539$). Table 1 shows σ_a^2 for each condition, and there were significant differences between the values of *Accurate* and *Fast* ($p < 0.01$) and *Neutral* and *Fast* ($p < 0.01$). According to the dual Gaussian distribution model, σ_a is not affected by the speed-accuracy tradeoff [10], but Yamanaka and Usuba reported that σ_a was significantly changed depending on the computation method (either a calibration task or a regression intercept) and the instruction in the calibration task (*Fast* or *Neutral*) [61]. Our results empirically support Yamanaka and Usuba's report.

4.1.2 Model Performance. We checked the radicand in the target size of Bi et al.'s FFitts law ($4.133\sqrt{\sigma^2 - \sigma_a^2}$). When σ_a was obtained from the calibration task, the radicand became negative in 9 / 18 conditions. Using σ_a from the intercept of $\sigma^2 = \text{constant} \times W^2 + \sigma_a^2$ made the radicand become negative in 2 / 18 conditions (Table 1). For these results, we could no longer use the original FFitts law (Eq. 5), aligning with a previous replication study [61]. Thus, in the 1D task, we verified three models (Fitts' law, Ko et al.'s FFitts law, and no-root Ko et al.'s FFitts law) for each of W and W_e; six candidate formulations in total.

To balance the model fit and complexity, we used adjusted R^2 and *AIC* [2] as evaluation metrics for all fitting points, i.e., *MT*s under three biases were regressed together. The model fit is better

Table 2. 1D: (a) Adjusted R^2 and *AIC* for all fitting points. (b) R^2, *MAE*, and *RMSE* for LOOCV.

	Model	*ID*	(a) All fitting points		(b) LOOCV		
			Adj. R^2	*AIC*	R^2	*MAE*	*RMSE*
#1	Fitts' law with W	$\log_2\left(\frac{A}{W}+1\right)$	0.133	238	0.0222	153	182
#2	Fitts' law with W_e	$\log_2\left(\frac{A}{W_e}+1\right)$	0.619	223	0.563	101	117
#3	Ko et al.'s FFitts law with W	$\log_2\left(\frac{A}{\sqrt{W^2-c^2}}+1\right)$	0.0851	239	0.00507	160	190
#4	Ko et al.'s FFitts law with W_e	$\log_2\left(\frac{A}{\sqrt{W_e^2-c^2}}+1\right)$	0.890	201	0.875	56.8	62.5
#5	No-root Ko et al.'s FFitts law with W	$\log_2\left(\frac{A}{W-c}+1\right)$	0.0853	239	0.00220	162	192
#6	No-root Ko et al.'s FFitts law with W_e	$\log_2\left(\frac{A}{W_e-c}+1\right)$	0.931	193	0.897	47.5	56.7

for higher adjusted R^2 and lower *AIC*. In addition, to evaluate the prediction accuracy for untested task conditions, we compared R^2, *MAE* (mean absolute error), and *RMSE* (root mean squared error) for the leave-one ($A \times W \times Bias$)-out cross-validation (LOOCV).

Table 2 summarizes the model fits for the six candidate formulations. Model #6 (no-root Ko et al.'s FFitts law with W_e) showed the best scores for all fitting points and LOOCV. Comparing the W and W_e versions for each model, using W_e instead of W improved scores, as it could normalize biases. This was also supported by the *TP*s shown in Fig. 2, where models using W_e yielded more stable *TP*s and thus their *TP* differences were smaller than when using W.

When calculating the fit and *TP* of Ko et al.'s formulations (Models #3–#6), the *ID* is determined through a nonlinear regression using the average *MT* from all participants. Consequently, individual participants do not have their own *TP*, which is why there are no error bars in Fig. 2. To align with this method, Fitts' law (Models #1–#2) also calculates *TP* using the average *MT* of all participants. There are multiple timings at which *TP* is computed (e.g., for each participant or for a group of participants [44]), and our approach is the one adopted in a previous study [62].

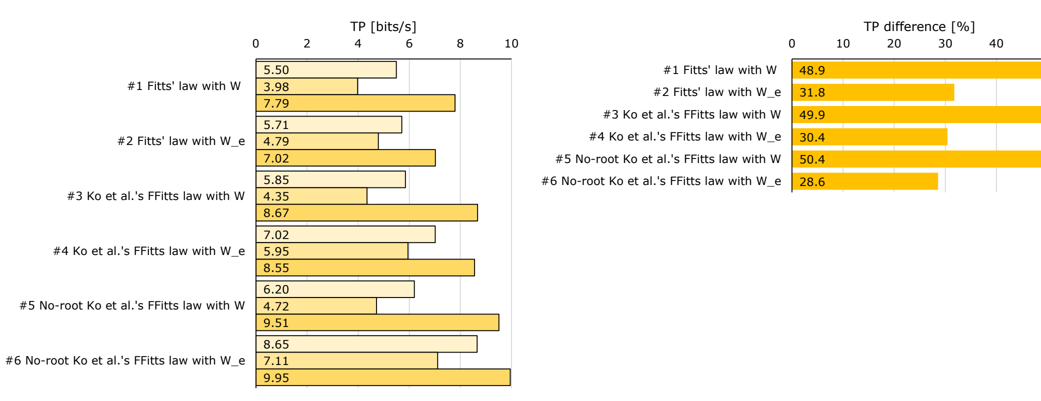


Fig. 2. 1D: (a) *TP* for each *Bias*. (b) *TP* difference.

Figure 3 shows the model fit for each *Bias*. In all models, using W showed better fits, particularly for Ko et al.'s FFitts law (Models #3 and #5). Thus, when researchers want to predict *MT* under a single instruction (typically *Neutral*), using W is better even if the participants' biases are shifted towards speed or accuracy. These results are consistent with those of previous studies [61–63].

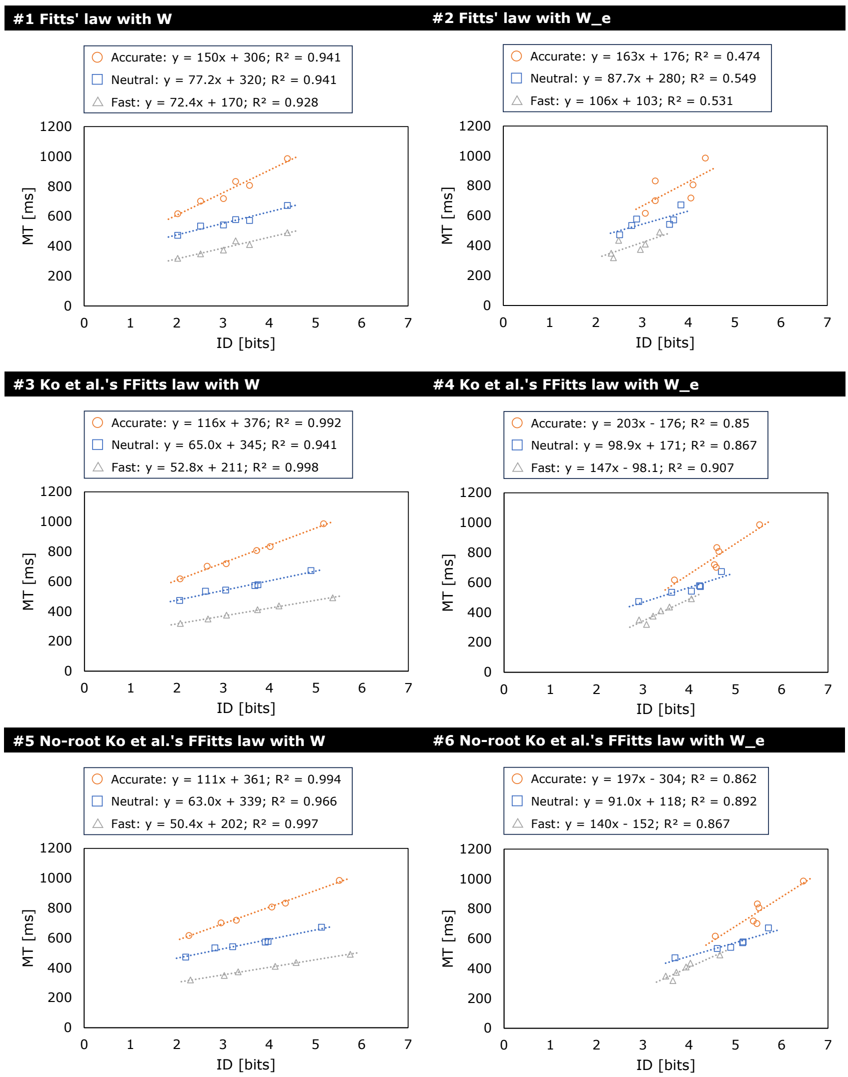


Fig. 3. 1D: Model fitting for each *Bias*.

4.2 2D Task

4.2.1 Movement Time, Error Rate, and σ_a. We again ran repeated-measures ANOVAs. Regarding *MT*, we found the main effects for *Bias* ($F_{1.33,19.9} = 54.3$, $p < 0.001$, $\eta_p^2 = 0.784$): 621 ms for *Neutral*, 860 ms for *Accurate*, and 443 ms for *Fast*, and all pairwise comparisons were significant ($p < 0.001$). Regarding *ER*, we found the main effects for *Bias* ($F_{1.44,21.6} = 22.7$, $p < 0.001$, $\eta_p^2 = 0.602$): 13.1% for *Neutral*, 7.64% for *Accurate*, and 21.4% for *Fast*, and all pairwise comparisons were significant ($p < 0.01$). Regarding σ_a^2, we found the main effect for *Bias* ($F_{1.38,20.6} = 7.50$, $p < 0.01$, $\eta_p^2 = 0.333$; see Table 2 for each value), and there were significant differences between *Accurate* and *Fast* ($p < 0.05$). Consistently with the 1D task, the participants appropriately followed each *Bias* instruction.

Table 3. 2D: $\sigma^2 - \sigma_a^2$ and $\sigma^2 - intercept$ for each condition. σ_a is obtained from the calibration task. *intercept* is obtained from the regression intercept. The red values are negative, and the target size in FFitts law is undefined.

		Neutral $\sigma_a^2 = 1.78$, $intercept = 2.03$			*Accurate* $\sigma_a^2 = 1.54$, $intercept = 1.60$			*Fast* $\sigma_a^2 = 2.16$, $intercept = 3.17$		
A	W	σ^2	$\sigma^2 - \sigma_a^2$	$\sigma^2 - intercept$	σ^2	$\sigma^2 - \sigma_a^2$	$\sigma^2 - intercept$	σ^2	$\sigma^2 - \sigma_a^2$	$\sigma^2 - intercept$
26	3	2.18	0.397	0.146	1.47	−0.0706	−0.122	2.85	0.684	−0.324
26	5.5	2.23	0.445	0.194	1.81	0.264	0.213	3.71	1.55	0.542
26	8.5	2.91	1.13	0.876	2.45	0.906	0.855	6.45	4.29	3.28
60	3	2.09	0.310	0.0591	1.80	0.253	0.202	4.10	1.93	0.927
60	5.5	3.00	1.22	0.969	2.03	0.481	0.430	5.66	3.50	2.49
60	8.5	3.54	1.75	1.50	1.96	0.416	0.365	6.29	4.12	3.12

Table 4. 2D: (a) Adjusted R^2 and *AIC* for all fitting points. (b) R^2, *MAE*, and *RMSE* for LOOCV.

			(a) All fitting points		(b) LOOCV		
	Model	*ID*	Adj. R^2	*AIC*	R^2	*MAE*	*RMSE*
#1	Fitts' law with W	$\log_2\left(\frac{A}{W}+1\right)$	0.192	242	0.0756	176	206
#2	Fitts' law with W_e	$\log_2\left(\frac{A}{W_e}+1\right)$	0.287	240	0.197	153	189
#3	Ko et al.'s FFitts law with W	$\log_2\left(\frac{A}{\sqrt{W^2-c^2}}+1\right)$	0.214	243	0.0694	178	212
#4	Ko et al.'s FFitts law with W_e	$\log_2\left(\frac{A}{\sqrt{W_e^2-c^2}}+1\right)$	0.718	224	0.688	97	117
#5	No-root Ko et al.'s FFitts law with W	$\log_2\left(\frac{A}{W-c}+1\right)$	0.217	243	0.0719	177	212
#6	No-root Ko et al.'s FFitts law with W_e	$\log_2\left(\frac{A}{W_e-c}+1\right)$	0.856	212	0.820	69.5	89.7

4.2.2 Model Performance. When σ_a was obtained from the calibration task, the radicand in Bi et al.'s FFitts law became negative in 1 / 18 conditions (Table 3). Using σ_a from the intercept method made the radicand negative in 2 / 18 cases. This prevented us from using FFitts law with σ_a, so we examined the same six models as in the 1D task.

Table 4 summarizes the model fits. Model #6 (no-root Ko et al.'s FFitts law with W_e) again showed the best scores for all fitting points and LOOCV, but the adjusted $R^2 = 0.856$ was somewhat lower than in the 1D task (0.931). The other results were consistent with the 1D task: using W_e instead of W improved the model-fit scores and the *TP* difference (Fig. 4), and using W showed a higher R^2 for each *Bias* (Fig. 5; using no-root Ko et al.'s FFitts law (Model #5) is the best).

5 Reanalysis of Batmaz et al.'s Datasets on Pointing Tasks in VR

To test our hypothesis that combining FFitts law with W_e can normalize speed-accuracy biases in broader conditions, we reanalyzed Batmaz et al.'s datasets [8]. In their experiment, spherical targets in a VR environment were successively selected with three bias conditions. In VR pointing tasks, fine control of selection position is challenging due to issues with a user's hand tremor [4, 40] and the input device's accuracy [33, 37, 38]. Considering that Ko et al.'s formulation (FFitts law with a free parameter) with W_e was originally proposed by Welford to absorb the impact of hand tremor in stylus-based pointing [53], and was later applied to mouse- [13] and finger-based pointing [35, 61], we expect this model to be applicable to VR pointing, where hand/controller tremor is more likely significant.

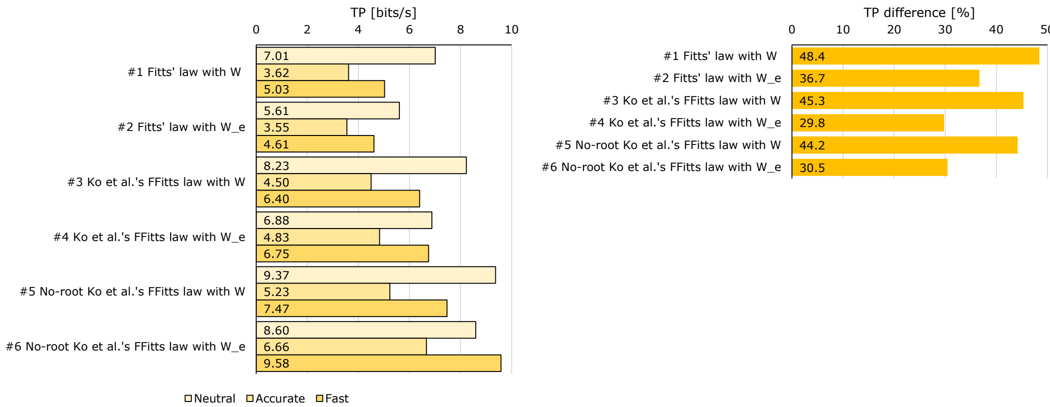


Fig. 4. 2D: (a) *TP* for each *Bias*. (b) *TP* difference.

5.1 Overview of Task, Procedure, and Performance Results

Eighteen participants wore an HTC VIVE Pro head-mounted display and selected targets by pressing a button on the VIVE hand-held controller. The experiment involved three types of bias (*Speed*, *None*, and *Precision*), two selection methods (ray casting and virtual hand), three *A*s (12.5, 25.0, and 37.5 cm), and three *W*s (sphere diameters; 1.5, 2.5, and 3.5 cm). Click-coordinate variabilities were measured as the univariate SD_x along the task axis.

In ray casting, the controller emits a ray that selects a target it crosses. For virtual hand, a dot fixed near the tip of the controller acts as the cursor, which requires larger arm movements compared to ray casting.

The speed-accuracy bias was adjusted by changing the pitch of the auditory feedback. Its design was inspired by a previous study [6] in which the pitch of the sound played for error clicking significantly affected the participants' subsequent speed-accuracy bias. On the basis of this result, the three feedback types for the current datasets are as follows.

- Speed-based feedback (*Speed*): During a trial, a sound is continuously played, and its pitch gradually increases. Thus, participants would naturally tend to operate faster to avoid hearing high (irritating) pitches.
- No auditory feedback (*None*).
- Precision-based feedback (*Precision*): A sound is played when clicking. The farther the click coordinate is from the target center, the higher the pitch. Participants would tend to operate more accurately to avoid high pitches.

Note that the participants are asked to perform the task as fast and accurately as possible. Hence, the auditory feedback was provided with the expectation that participants would implicitly change their speed-accuracy bias.

For ray casting, the results showed that the *MT*s for the *Speed*, *None*, and *Precision* conditions were 710, 1017, and 1201 ms, respectively, and the *ER*s were 10.9, 10.1, and 2.61%. Similarly, for virtual hand, the *MT*s were 699, 992, and 1131 ms, and *ER*s were 6.80, 6.56, and 1.17%. Hence, the feedback for the *None*, *Speed*, and *Precision* conditions had the intended effect to change the speed-accuracy bias. For more detailed experimental designs and sound pitch frequencies, please refer to the original paper [8].

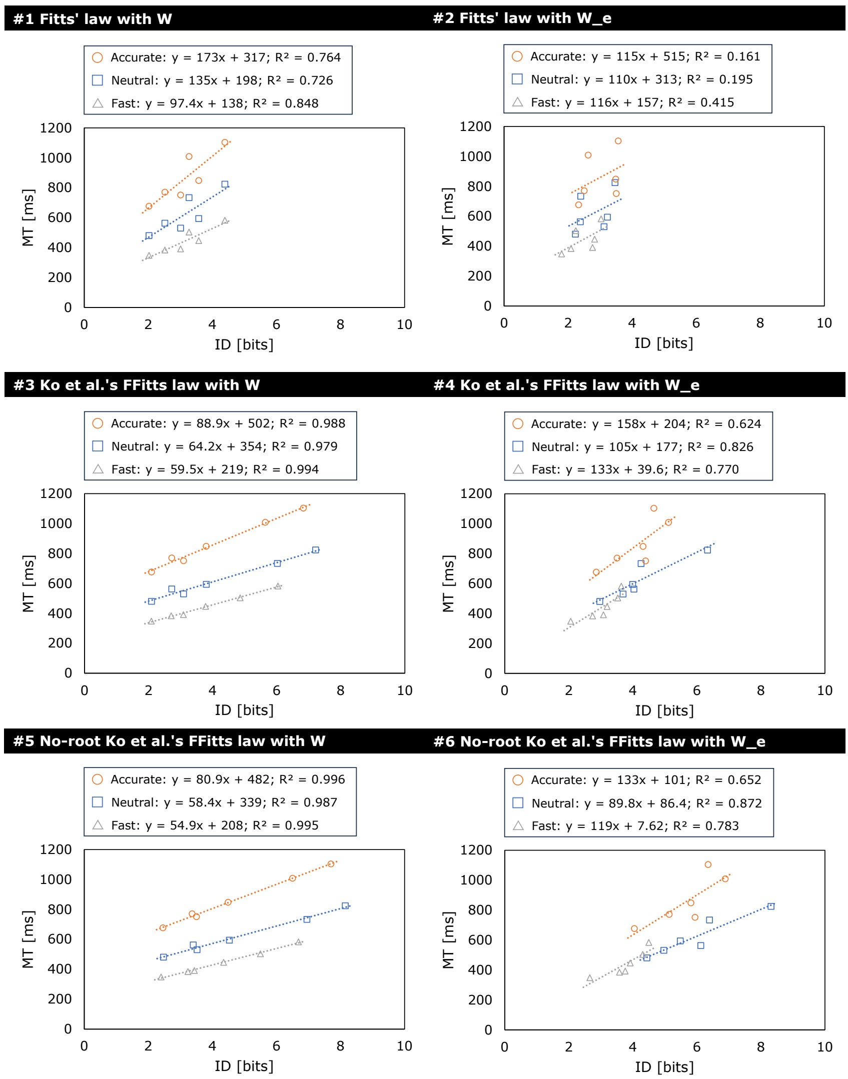


Fig. 5. 2D: Model fitting for each *Bias*.

5.2 Model Performance

Because the resultant effects of the auditory feedback on *MT* corresponded to *Fast*, *Neutral*, and *Accurate* in our experiment, we hereafter use these three notations for consistency of terminology. Batmaz et al.'s experiment does not include a calibration task, and thus we directly analyze model fits of the six formulations and the stability of *TP*.

5.2.1 Ray Casting. Table 5 summarizes the model fits. Model #6 showed the best scores for all fitting points and LOOCV. Using W_e instead of W improved scores for all three models. Figure 6 shows that Model #6 again yielded the most stable *TP*s across the three bias conditions.

Table 5. Batmaz et al., ray casting: (a) adjusted R^2 and *AIC* for all fitting points and (b) R^2, *MAE*, and *RMSE* for LOOCV.

	Model	*ID*	(a) All fitting points		(b) LOOCV		
			Adj. R^2	*AIC*	R^2	*MAE*	*RMSE*
#1	Fitts' law with W	$\log_2\left(\frac{A}{W}+1\right)$	0.244	369	0.164	197	233
#2	Fitts' law with W_e	$\log_2\left(\frac{A}{W_e}+1\right)$	0.501	358	0.429	153	183
#3	Ko et al.'s FFitts law with W	$\log_2\left(\frac{A}{\sqrt{W^2-c^2}}+1\right)$	0.228	370	0.150	198	226
#4	Ko et al.'s FFitts law with W_e	$\log_2\left(\frac{A}{\sqrt{W_e^2-c^2}}+1\right)$	0.700	346	0.624	132	151
#5	No-root Ko et al.'s FFitts law with W	$\log_2\left(\frac{A}{W-c}+1\right)$	0.228	370	0.127	201	231
#6	No-root Ko et al.'s FFitts law with W_e	$\log_2\left(\frac{A}{W_e-c}+1\right)$	0.713	343	0.681	120	137

Figure 7 shows the model fitting for each *Bias*. In all models, using W showed better fits than using W_e for each *Bias*. Ko et al.'s FFitts law (Models #3 and #5) showed almost equally the best R^2 values. These results are consistent with our results in the 1D and 2D tasks.

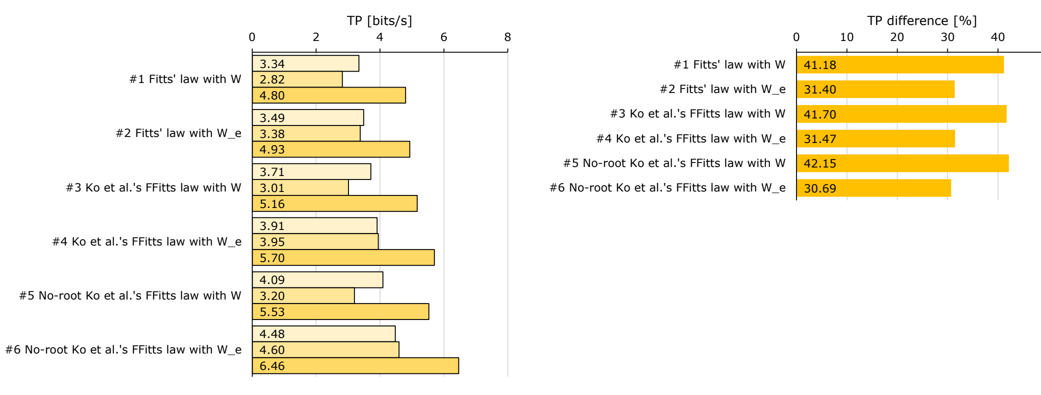


Fig. 6. Batmaz et al., ray casting: (a) *TP* for each *Bias* and (b) *TP* difference

.

5.2.2 *Virtual Hand.* Table 6 indicates consistent results with ray casting: Model #6 showed the best scores for all fitting points and LOOCV, and using W_e showed better scores. However, Fig. 8 shows that Model #2 (Fitts' law with W_e) yielded the smallest *TP* difference of 28.47%. This is a unique result among the four datasets we examined.

Figure 9 indicates that using W showed better fits than W_e for each *Bias*. The baseline Fitts' law (Model #1) and two models of Ko et al.'s FFitts law (Models #3 and #5) showed similarly high R^2 values. The R^2 values remained high ($0.918 < R^2 < 0.993$) compared to the ray-casting data ($0.662 < R^2 < 0.960$ in Fig. 7).

6 Discussion

6.1 Model Performance in Touch-Pointing Experiment

We showed that using W_e improved the model fit for all fitting points and decreased the *TP* difference, while using W yielded a higher R^2 for each *Bias*. These effects have already been confirmed in pointing tasks with pen tablets [63] and mice [62], and we showed that these also

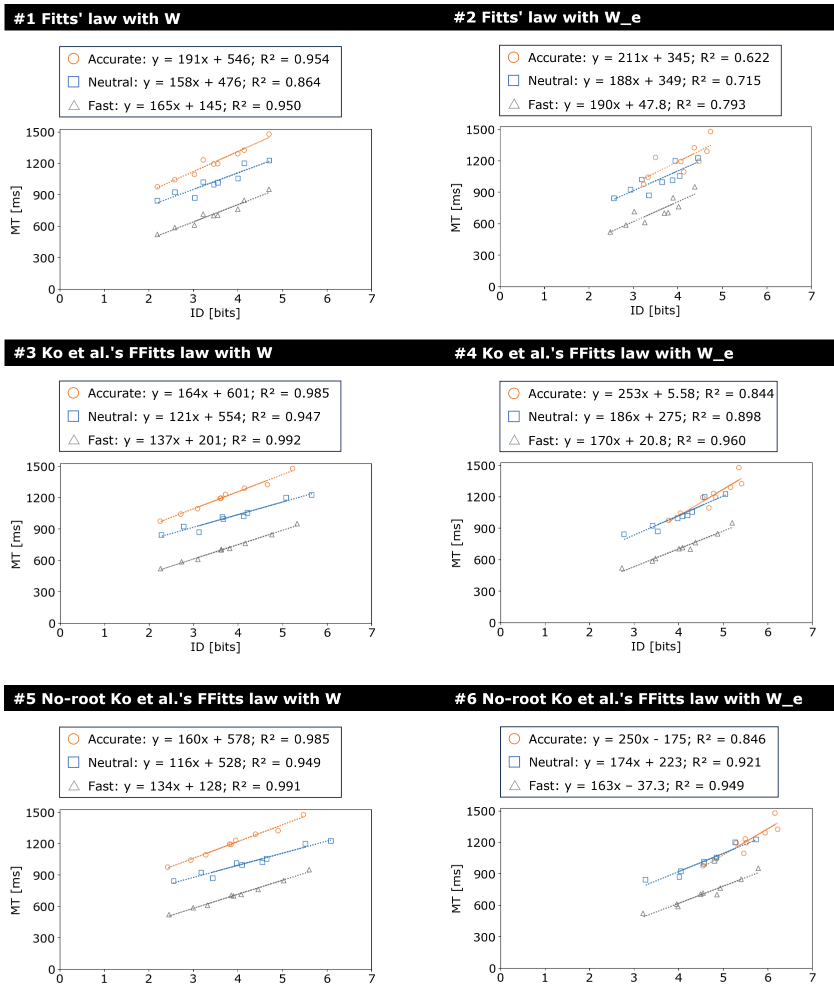


Fig. 7. Batmaz et al., ray casting: model fitting for each *Bias*.

apply to finger-touch pointing. Thus, the advantages and disadvantages of using W_e shown in our tasks are consistent with those in previous studies.

In both 1D and 2D tasks, we could not use FFitts laws with σ_a in some conditions because of negative radicands ($\sigma^2 - \sigma_a^2 < 0$; i.e., $\sigma < \sigma_a$). This was more evident in *Accurate* and *Fast* than in *Neutral*. In the 1D task, σ_a^2 was 1.01 in *Neutral* and 0.814 in *Accurate* (Table 1); i.e., a decrease of 19.4% from *Neutral* to *Accurate*. On the other hand, σ decreased by 51.9% on average across the six target conditions, which indicates that the degree to which σ was reduced is more substantial than that for σ_a. This could presumably be why we obtained the negative radicands more frequently under the *Accurate* condition than *Neutral*.

Bi et al. assumed that the speed-accuracy bias affects only the relative term σ_r in Eq. 4 ($\sigma^2 = \sigma_r^2 + \sigma_a^2$) [10]. However, the results of our two tasks contradicted this assumption, and supported Yamanaka and Usuba's report on the significant effect of *Bias* on σ_a [61]. We thus agree with their suggestion that σ_a should not be used.

Table 6. Batmaz et al., virtual hand: (a) adjusted R^2 and *AIC* for all fitting points and (b) R^2, *MAE*, and *RMSE* for LOOCV.

	Model	*ID*	(a) All fitting points		(b) LOOCV		
			Adj. R^2	*AIC*	R^2	*MAE*	*RMSE*
#1	Fitts' law with W	$\log_2\left(\frac{A}{W}+1\right)$	0.398	362	0.328	175	198
#2	Fitts' law with W_e	$\log_2\left(\frac{A}{W_e}+1\right)$	0.687	344	0.645	123	143
#3	Ko et al.'s FFitts law with W	$\log_2\left(\frac{A}{\sqrt{W^2-c^2}}+1\right)$	0.372	364	0.307	178	202
#4	Ko et al.'s FFitts law with W_e	$\log_2\left(\frac{A}{\sqrt{W_e^2-c^2}}+1\right)$	0.755	341	0.702	115	132
#5	No-root Ko et al.'s FFitts law with W	$\log_2\left(\frac{A}{W-c}+1\right)$	0.372	364	0.304	178	202
#6	No-root Ko et al.'s FFitts law with W_e	$\log_2\left(\frac{A}{W_e-c}+1\right)$	0.762	340	0.709	112	130

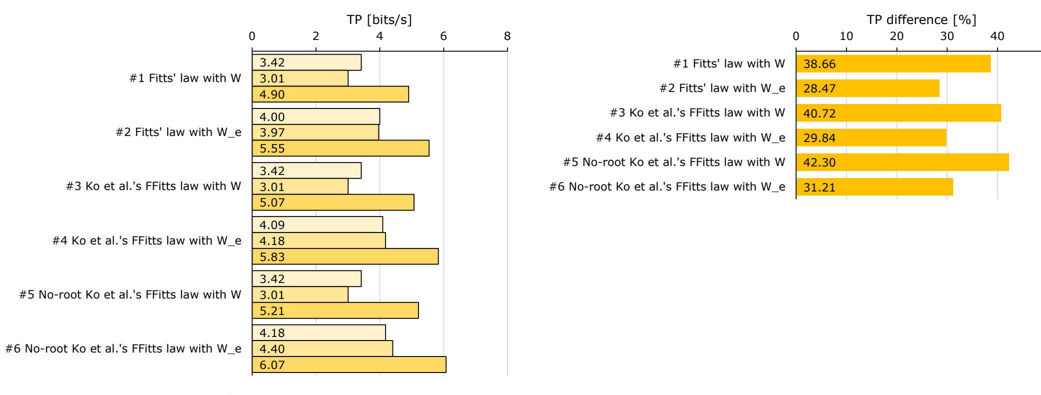


Fig. 8. Batmaz et al., virtual hand: (a) *TP* for each *Bias* and (b) *TP* difference

.

Of course, there is a chance that researchers might obtain no negative radicands. For example, in our 2D task, if we had not tested the condition of $A = 26$ and $W = 3$ mm, we could have used FFitts law due to all radicands being positive (Table 3). However, σ and σ_a are random variables that differ depending on the target conditions and input devices. Therefore, to analyze Fitts-task data systematically, we recommend applying Ko et al.'s method (using c instead of σ_a). Moreover, a calibration task takes extra time in addition to that required for a Fitts task, so as a practical implication for future researchers, we recommend *not* conducting it.

In Yamanaka and Usuba's report, the best model to predict *MT* depends on datasets. However, in our study, the no-root Ko et al.'s FFitts law with W_e (Model #6) always showed the best scores in all data points and LOOCV in both the 1D and 2D tasks. Thus, if researchers conduct touch-pointing studies to compare several experimental factors, we recommend using Model #6 for fair comparison. In conclusion, for touch-pointing tasks, our hypothesis is supported: combining the effective width method with FFitts law is beneficial to normalize subjective biases.

6.2 Reasons behind Poor Fits of Baseline Models in the 2D Task

The results of our 2D task indicate that even under the common *Neutral* condition, the fit of Fitts' law with W (Model #1) and that with W_e (#2) was low: R^2 values were 0.726 and 0.195, respectively (Fig. 5). In comparison, previous studies using 2D circular targets with *Neutral* instruction have

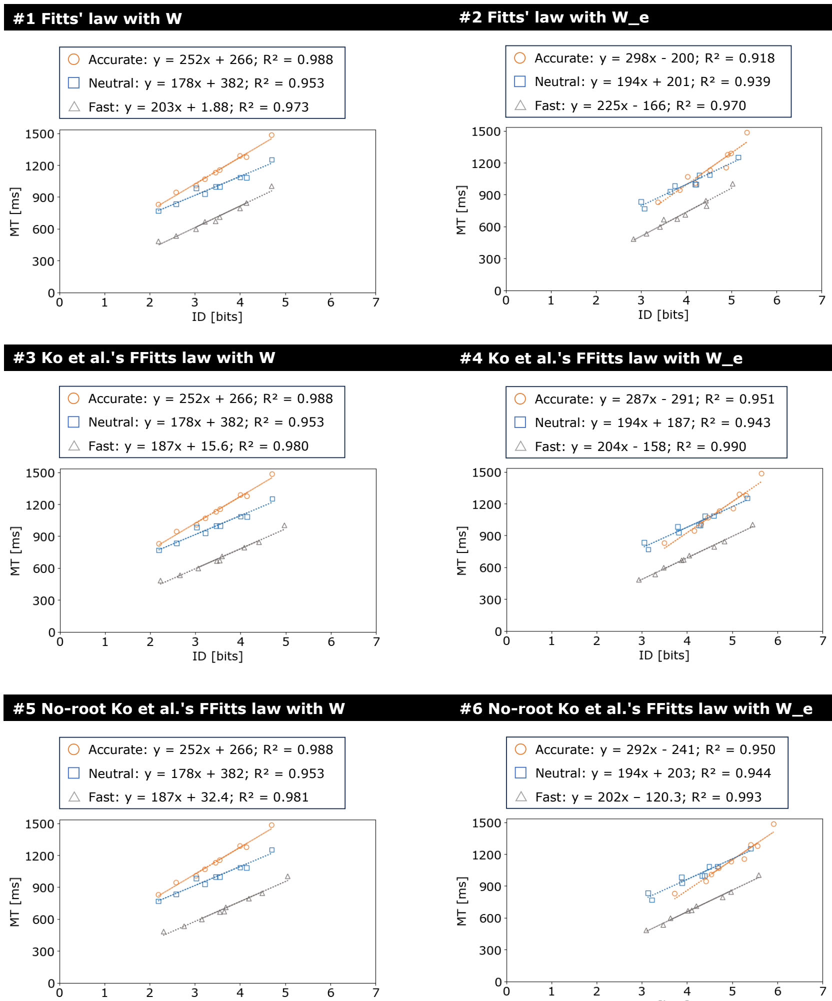
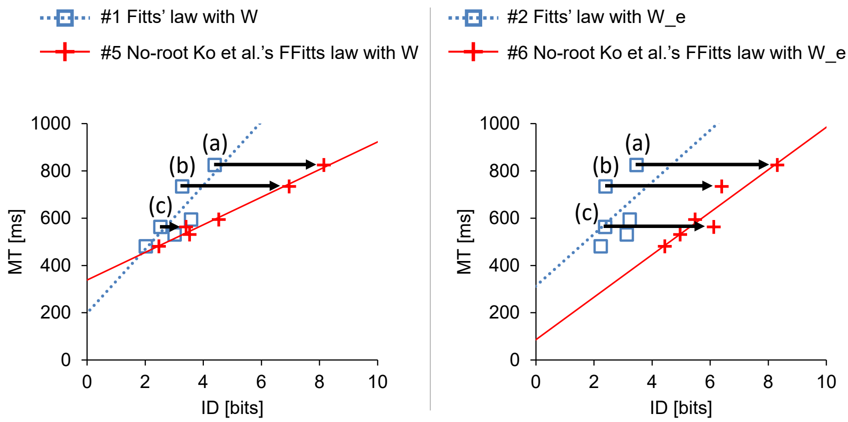


Fig. 9. Batmaz et al., virtual hand: model fitting for each *Bias*.

shown better fits. For example, in the experiment by Yamanaka and Usuba, Model #1 had an R^2 of 0.9904, and #2 had an R^2 of 0.7317 [61]. Bi et al. also showed R^2 values of 0.85 and 0.79, respectively [10]. We discuss possible reasons for the poor fits in our results below.

(1) Some target conditions could be completed with ballistic movements. According to Hoffmann et al., when Fitts' original $ID = \log_2(2A/W)$ is less than 3–4 bits, the pointing task can be completed without visual feedback [26]. Such ballistic movements are done in less than the human corrective reaction time (~260 ms), and do not fit well with Fitts' law. In our experiment, three conditions had *ID*s below 4 bits: (A, W, Fitts' ID) = (26, 5.5, 3.24), (26, 8.5, 2.61), (60, 8.5, 3.82). However, looking at Model #1 in Fig. 5, because *MT*s were longer than 400 ms under the *Neutral* condition, the movements cannot be considered completely ballistic, and it does not seem to be a definitive cause of poor model fit.

(2) The numbers of participants and trials were not sufficient. Because the original Fitts' law (Model #1) is to predict the average *MT* by numerous participants, the model fit should improve



Data point	A [mm]	W [mm]	ID (#1) [bits]	ID (#5) [bits]	We [mm]	ID (#2) [bits]	ID (#6) [bits]	MT [ms]	ER [%]
(a)	60	3	4.392	8.151	5.975	3.465	8.313	824	33.20
(b)	26	3	3.273	6.951	6.099	2.396	6.394	734	35.10
(c)	26	5.5	2.518	3.404	6.165	2.383	6.121	563	1.953

Fig. 10. The results from our 2D task to focus on the poor fit of Models #1 and #2 under the *Neutral* condition. Black arrows indicate how FFitts law moves the data points by adjusting the *ID*s.

as the numbers of participants and trials increase. Also, Fitts' law with W_e (Model #2) is derived on the basis that the more tapping actions are performed, the closer the endpoints will be to a normal distribution. Previous studies have experimentally shown that increasing the numbers of participants and trials improve R^2 of Models #1 and #2 [58, 59]. However, the aforementioned experiments by Bi et al. and Yamanaka and Usuba recruited 12 participants performing 16 repetitions, while we employed 16 participants and 16 repetitions. Therefore, this also does not seem to be a definitive cause for the poor fit.

(3) Some task conditions were too difficult. Experiments that fit Fitts' law typically have an *ER* of 4–5% [47]. In our 2D experiment, even under the *Neutral* condition, the average *ER* was 13.1%. Thus, *MT*s with smooth operations might not have been measured, which potentially disrupted the relationship modeled by Fitts' law. Particularly for the target with $W = 3$ mm, the *ER* exceeded 30% (see Fig. 10), and the relatively long time compared to other *W*s might have worsened the model fit. However, even in Bi et al.'s experiment [10], where *W*s of 2.4, 4.8, and 7.2 mm were used with an average *ER* of 33%, the fit was high despite more difficult conditions than in our experiment.

(4) Impact of the fat finger problem was greater for 2D targets than 1D. The phenomenon that a finger occludes a target is called the fat finger problem [29, 52]. In 1D tasks, because the targets were horizontally long, it was relatively easy to adjust the finger position on the target even if it was small. However, for small 2D targets, the finger completely occludes the target just tapping, which increases the necessity of careful finger positioning. According to Fig. 10a and b, the two conditions of $W = 3$ mm exhibited *ER*s over 33% and had the highest *MT*s among the six target conditions. This indicates that much higher *ID*s should be used for these data points, but Fitts' law with *W* (Model #1) yielded *ID* values close to those of other target conditions, while Model #5 resolved this issue.

Regarding Fitts' law with W_e (Model #2), for small 2D targets, careful operation did not necessarily result in tapping near the target. According to Fig. 10b and c, for $W = 3$ and 5.5 mm, the *MT*s were

734 and 563 ms, respectively (a difference of 23.3%), but the W_e values are 6.099 and 6.165 mm (a difference of only 1.08%). These results led to a large difference on the y-axis values of (b) and (c) in Fig. 10 right panel, while there is almost no difference on the x-axis values, which could have result in the poor fit.

(5) A tablet and discrete pointing task were used. The experiments by Bi et al. [10] and Yamanaka and Usuba [61] were conducted on smartphones, and a set of the start button and target was presented for each trial, while our design involved tapping targets successively. The former is called a discrete Fitts task while the latter is known as a serial task [24, 56]. In the case of discrete tasks, it might be possible to sufficiently locate the target position before tapping the start button in each trial, and perhaps tilt the smartphone held in the non-dominant hand to peek at the target if necessary. In contrast, we used a tablet placed on a table, and thus it would be challenging to peek at each target, and doing so would increase the *MT*. This might disrupt the lawful regularity in 2D tasks.

6.3 Model Performance in Batmaz et al.'s Experiment

Batmaz et al.'s instructions were to perform the task as fast and accurately as possible, but auditory feedback led the participants to change their speed-accuracy bias. Our reanalysis showed that, at least for ray casting, using the effective width method and FFitts law together appropriately normalized the bias.

The model fits for all fitting points were worse compared to our touch-pointing experiment; the best adjusted R^2 for ray casting was 0.713 and that for virtual hand was 0.762. Previous studies have identified several factors influencing why operations in VR encounter difficulties and affecting Fitts' law fits, such as stereoscopic vision [7, 51] and the need to keep one's arm raised [25, 30].

Since the type of auditory feedback significantly influenced *MT*, using Fitts' law with *W* (Model #1) made the three regression lines diverge vertically (Figs. 7a and 9a). Therefore, when predicting *MT* data while mixing the three bias conditions, using W_e instead of *W* provided a better model fit for both selection methods (Tables 5 and 6). In particular, Model #6 showed the best fit, and also had the highest capability to normalize *TP* in ray casting.

The exception was the stability of *TP* in virtual hand, where Fitts' law with W_e was the best (Model #2). One reason for this could be the lesser presence of tremor compared to ray casting. Even minor hand movements significantly shift the ray's endpoint in ray casting, whereas finer positioning was possible in virtual hand. This could reduce the contribution of the free parameter *c* for absorbing hand tremor. Thus, using the effective width method alone was sufficient to normalize *TP*, which made Model #2 the best. However, Table 6 showed that Model #6 was superior in model fit for virtual hand, so from this perspective, we recommend using Model #6. Thus, for virtual hand, it is logical to withhold a conclusion about the best model for normalizing the bias. In conclusion, our hypothesis on the benefit of combining the effective width method with FFitts law is supported only for ray casting.

6.4 Implications

6.4.1 Recommended Model for Comparative Study. To calculate *TP*s in touch-pointing tasks, researchers have used Fitts' law with *W* (Model #1) for comparing the effect of the participants' ages [57], or Fitts' law with W_e (Model #2) for comparing crowdsourced vs. laboratory participants [18]. Similarly, in VR pointing tasks with ray casting, Model #2 has been used. For example, a ray of infinite length yielded a significantly better *TP* than a fixed-length ray [5]. Gabel et al. also used Model #2 to compute *TP*s for comparing the performance of straight, curved, and rotated rays [21].

However, according to the results of our experiment and reanalysis, if the participants in the above four studies [5, 18, 21, 57] had an implicit bias towards speed or accuracy, it could prevent fair comparison. In particular, our *TP* comparison (e.g., Fig. 6a) showed that there is a tendency for *TP* to be calculated as higher when the bias is towards *Fast*. To avoid this effect, we recommend that researchers use Model #6 when comparing performance in touch operations and ray casting in their future studies.

6.4.2 Potential Applicability to Broader Experimental Situations. We limited the experimental data in this study to basic touch pointing on a PC display and pointing in VR, but there are other operational scenarios that involve decreased touch precision. Previous studies have reported that pointing performance deteriorates (e.g., *MT* and *ER* increase) with intensified vibration, such as operating touchscreens in car navigation systems [1, 48] or in airplane cockpits [15, 16, 50]. Additionally, the accuracy of touch operations on smartphones decreases with faster walking speed [9, 39, 46].

In these papers, *MT* and *ER* were compared independently, but using a single metric of *TP* can simplify comparisons. For example, in Tao et al.'s study on comparing touchscreen performance in cars at different vibration levels, *MT* and *ER* were reported separately [49]. Figure 2 of their paper shows that, although the participants were instructed to maintain a 1% *ER* across three vibration levels (Static, Slight, and Moderate), the smallest *MT* was observed for the Slight condition. This is contrary to other vibration-related previous studies where the Static condition exhibits the smallest *MT* [1, 16, 46]. Therefore, there could be an implicit bias towards speed under the Slight condition.

For fairer comparison between task conditions, computing *TP*s after normalizing such biases can facilitate discussions about performance changing. Based on our experimental results, for future researchers who would like to conduct comparative experiments, we recommend using a model that combines FFitts law and the effective width method.

6.5 Limitations and Future Work

In our 2D task, while the no-root Ko et al.'s FFitts law with W_e (Model #6) showed the best model-fit scores, the adjusted $R^2 = 0.856$ was somewhat lower than in 1D. However, it still remained high compared to previous studies, e.g., $R^2 = 0.825$ for three biases and $R^2 = 0.783$ for five biases using stylus [63]. Note that a direct comparison is difficult due to differences in tasks, and also that deciding a model's goodness of fit based solely on R^2 is inadequate [22]. Thus, we focused on comparatively discussing the relative highs and lows across different models and datasets.

Using the circular shape for 2D targets is another limitation. In mouse-based rectangular target pointing with three biases, a model using W_e and the effective height H_e showed better fit scores than using only W_e [62]. Thus, our next step is to verify whether combining FFitts law with W_e and H_e can normalize the biases for rectangular targets.

We expected that a single model would consistently show the best performance in terms of fit scores and *TP*-stabilizing effects in all of the 1D and 2D touch-pointing tasks, as well as the two VR pointing tasks. However, although Model #6 often normalized the biases the most, Fitts' law with W_e (Model #2) showed the best result for *TP* stability in the virtual hand condition. Investigating a model that can robustly achieve the highest bias-normalization capability under much broader experimental conditions will be included in our future work.

7 Conclusion

In this study, we investigated FFitts law under subjective biases for pointing tasks with a touchscreen and a VR environment. Using W_e instead of W improved the model fit when the data combined three bias conditions and narrowed the *TP* range. In comparison, analyzing the *MT* data for a single bias showed that using W enabled a better fit. These results are consistent with mouse pointing. In addition, for touch operations, we found that $\sigma^2 - \sigma_a^2$ becomes negative in some conditions, and the no-root Ko et al.'s FFitts law with W_e (Model #6) showed the best scores in the 1D and 2D tasks, as well as the ray-casting operation in VR. Thus, if researchers conduct experiments with the subjective biases in finger pointing and ray casting, we recommend using Model #6, while using either of the two formulations of Ko et al.'s FFitts law with W (Models #3 and #5) was often the best for predicting MT under a single bias.

In the HCI field, researchers proposing novel touch-input devices or target-acquisition techniques have compared their methods with a certain baseline, such as common touchscreens or unassisted tap operations. In such cases, simply comparing MTs might lead participants to implicitly prioritize speed or accuracy, and thus researchers have been suggested to use TP [41, 47]. Without our current paper, they would likely follow the standardized methodology, i.e., using the ISO 9241-9 task (where circular targets are arranged in a circle) and calculating TP by Fitts' law with W_e (Model #2).

However, as our experimental results showed, Model #2 does not sufficiently normalize the effects of bias, and thus participants may have a bias towards speed or accuracy, which may not allow for a fair comparison of input devices or interaction techniques. Additionally, simply combining three separate findings from previous research to conclude that (1) it is better to use W_e to calculate TP for device comparison, (2) it is better to use the Finger-Fitts law for touch operations, and (3) Ko et al.'s modified Finger-Fitts model is better than the original version, might lead some researchers to coincidentally use the best model (#6) by chance. However, this is scientifically inappropriate as it is not based on evidence. This paper provides crucial evidence on which models should be used in future research methodologies and offers insights on users' touch-operation behaviors to the audience within the MobileHCI community.

References

[1] Bashar I. Ahmad, Patrick M. Langdon, Simon J. Godsill, Robert Hardy, Lee Skrypchuk, and Richard Donkor. 2015. Touchscreen Usability and Input Performance in Vehicles under Different Road Conditions: An Evaluative Study. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Nottingham, United Kingdom) *(AutomotiveUI '15)*. Association for Computing Machinery, New York, NY, USA, 47–54. https://doi.org/10.1145/2799250.2799284

[2] 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

[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)*. Association for Computing Machinery, New York, NY, USA, 1837–1846. https://doi.org/10.1145/2702123.2702439

[4] Marc Baloup, Thomas Pietrzak, and Géry Casiez. 2019. RayCursor: A 3D Pointing Facilitation Technique Based on Raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300331

[5] Anil Ufuk Batmaz and Wolfgang Stuerzlinger. 2020. Effect of Fixed and Infinite Ray Length on Distal 3D Pointing in Virtual Reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3334480.3382796

[6] Anil Ufuk Batmaz and Wolfgang Stuerzlinger. 2021. Effects of Different Auditory Feedback Frequencies in Virtual Reality 3D Pointing Tasks. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, Washington, DC, USA, 189–194. https://doi.org/10.1109/VRW52623.2021.00042

[7] Anil Ufuk Batmaz, Rumeysa Turkmen, Mine Sarac, Mayra Donaji Barrera Machuca, and Wolfgang Stuerzlinger. 2023. Re-investigating the Effect of the Vergence-Accommodation Conflict on 3D Pointing. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology* (Christchurch, New Zealand) *(VRST '23)*. Association for Computing Machinery, New York, NY, USA, Article 8, 10 pages. https://doi.org/10.1145/3611659.3615686

[8] Anil Ufuk Batmaz, Kangyou Yu, Hai-Ning Liang, and Wolfgang Stuerzlinger. 2022. Improving Effective Throughput Performance using Auditory Feedback in Virtual Reality. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction* (Online, CA, USA) *(SUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 18, 11 pages. https://doi.org/10.1145/3565970.3567702

[9] Joanna Bergstrom-Lehtovirta, Antti Oulasvirta, and Stephen Brewster. 2011. The Effects of Walking Speed on Target Acquisition on a Touchscreen Interface. In *Proceedings of the 13th International Conference on Human Computer*

Interaction with Mobile Devices and Services (Stockholm, Sweden) *(MobileHCI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/2037373.2037396

[10] 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)*. Association for Computing Machinery, New York, NY, USA, 1363–1372. https://doi.org/10.1145/2470654.2466180

[11] Xiaojun Bi and Shumin Zhai. 2013. Bayesian Touch: A Statistical Criterion of Target Selection with Finger Touch. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (St. Andrews, Scotland, United Kingdom) *(UIST '13)*. Association for Computing Machinery, New York, NY, USA, 51–60. https://doi.org/10.1145/2501988.2502058

[12] Stuart K. Card, William K. English, and Betty J. Burr. 1978. Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys, and Text Keys for Text Selection on a CRT. *Ergonomics* 21, 8 (1978), 601–613. https://doi.org/10.1080/00140137808931762 arXiv:https://doi.org/10.1080/00140137808931762

[13] Olivier Chapuis and Pierre Dragicevic. 2011. Effects of Motor Scale, Visual Scale, and Quantization on Small Target Acquisition Difficulty. *ACM Trans. Comput.-Hum. Interact.* 18, 3, Article 13 (aug 2011), 32 pages. https://doi.org/10.1145/1993060.1993063

[14] Olivier Chapuis, Jean-Baptiste Labrune, and Emmanuel Pietriga. 2009. DynaSpot: Speed-dependent Area Cursor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. ACM, New York, NY, USA, 1391–1400. https://doi.org/10.1145/1518701.1518911

[15] Andy Cockburn, Carl Gutwin, Philippe Palanque, Yannick Deleris, Catherine Trask, Ashley Coveney, Marcus Yung, and Karon MacLean. 2017. *Turbulent Touch: Touchscreen Input for Cockpit Flight Displays.* Association for Computing Machinery, New York, NY, USA, 6742–6753. https://doi.org/10.1145/3025453.3025584

[16] Louise V. Coutts, Katherine L. Plant, Mark Smith, Luke Bolton, Katie J. Parnell, James Arnold, and Neville A. Stanton. 2019. Future technology on the flight deck: assessing the use of touchscreens in vibration environments. *Ergonomics* 62, 2 (2019), 286–304. https://doi.org/10.1080/00140139.2018.1552013

[17] Peter Dixon. 2008. Models of accuracy in repeated-measures designs. *Journal of Memory and Language* 59, 4 (2008), 447–456.

[18] Leah Findlater, Joan Zhang, Jon E. Froehlich, and Karyn Moffatt. 2017. Differences in Crowdsourced vs. Lab-Based Mobile and Desktop Input Performance Data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6813–6824. https://doi.org/10.1145/3025453.3025820

[19] 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

[20] Clifton Forlines and Ravin Balakrishnan. 2008. Evaluating Tactile Feedback and Direct vs. Indirect Stylus Input in Pointing and Crossing Selection Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. ACM, New York, NY, USA, 1563–1572. https://doi.org/10.1145/1357054.1357299

[21] Jenny Gabel, Susanne Schmidt, Oscar Ariza, and Frank Steinicke. 2023. Redirecting Rays: Evaluation of Assistive Raycasting Techniques in Virtual Reality. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology* (Christchurch, New Zealand) *(VRST '23)*. Association for Computing Machinery, New York, NY, USA, Article 38, 11 pages. https://doi.org/10.1145/3611659.3615716

[22] Julien Gori, Olivier Rioul, Yves Guiard, and Michel Beaudouin-Lafon. 2018. The Perils of Confounding Factors: How Fitts' Law Experiments can Lead to False Conclusions. 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, 1–10. https://doi.org/10.1145/3173574.3173770

[23] Tovi Grossman and Ravin Balakrishnan. 2005. The Bubble Cursor: Enhancing Target Acquisition by Dynamic Resizing of the Cursor's Activation Area. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. ACM, New York, NY, USA, 281–290. https://doi.org/10.1145/1054972.1055012

[24] Yves Guiard. 1997. Fitts' law in the discrete vs. cyclical paradigm. *Human Movement Science* 16, 1 (1997), 97–131. https://doi.org/10.1016/S0167-9457(96)00045-0

[25] Darren Guinness, Alvin Jude, G. Michael Poor, and Ashley Dover. 2015. Models for Rested Touchless Gestural Interaction. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction* (Los Angeles, California, USA) *(SUI '15)*. Association for Computing Machinery, New York, NY, USA, 34–43. https://doi.org/10.1145/2788940.2788948

[26] 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

[27] 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)*. Association for Computing Machinery, New York, NY, USA, 581–590. https://doi.org/10.1145/1753326.1753413

[28] 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)*. Association for Computing Machinery, New York, NY, USA, 2501–2510. https://doi.org/10.1145/1978942.1979308

[29] 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)*. Association for Computing Machinery, New York, NY, USA, 2501–2510. https://doi.org/10.1145/1978942.1979308

[30] Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. 2017. Modeling Cumulative Arm Fatigue in Mid-Air Interaction Based on Perceived Exertion and Kinetics of Arm Motion. 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, 3328–3339. https://doi.org/10.1145/3025453.3025523

[31] Izabelle Janzen, Vasanth K. Rajendran, and Kellogg S. Booth. 2016. Modeling the Impact of Depth on Pointing Performance. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 188–199. https://doi.org/10.1145/2858036.2858244

[32] Anirudha Joshi, Aaditya Ajikumar, Kuldeep Umaraiya, and Prashik Chavan. 2022. Fitts' Throughput and "Absolute" Finger Precision by Handedness, Hand, Digit and Target Width. In *Nordic Human-Computer Interaction Conference* (Aarhus, Denmark) *(NordiCHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 57, 11 pages. https://doi.org/10.1145/3546155.3546644

[33] Jari Kangas, Sriram Kishore Kumar, Helena Mehtonen, Jorma Järnstedt, and Roope Raisamo. 2022. Trade-Off between Task Accuracy, Task Completion Time and Naturalness for Direct Object Manipulation in Virtual Reality. *Multimodal Technologies and Interaction* 6, 1 (2022), 15 pages. https://doi.org/10.3390/mti6010006

[34] Nobuhito Kasahara, Yosuke Oba, Shota Yamanaka, Wolfgang Stuerzlinger, and Homei Miyashita. 2023. Throughput and Effective Parameters in Crossing. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 282, 9 pages. https://doi.org/10.1145/3544549.3585817

[35] Yu-Jung Ko, Hang Zhao, Yoonsang Kim, IV Ramakrishnan, Shumin Zhai, and Xiaojun Bi. 2020. Modeling Two Dimensional Touch Pointing. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 858–868. https://doi.org/10.1145/3379337.3415871

[36] Yu-Jung Ko, Hang Zhao, Iv Ramakrishnan, Shumin Zhai, and Xiaojun Bi. 2022. Issues Related to Using Finger-Fitts Law to Model One-Dimensional Touch Pointing Tasks. In *The Ninth International Symposium of Chinese CHI* (Online, Hong Kong) *(Chinese CHI 2021)*. Association for Computing Machinery, New York, NY, USA, 41–49. https://doi.org/10.1145/3490355.3490360

[37] Werner A. König, Jens Gerken, Stefan Dierdorf, and Harald Reiterer. 2009. Adaptive pointing: implicit gain adaptation for absolute pointing devices. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI EA '09)*. Association for Computing Machinery, New York, NY, USA, 4171–4176. https://doi.org/10.1145/1520340.1520635

[38] Sriram Kishore Kumar, Jari Kangas, Helena Mehtonen, Jorma Järnstedt, and R. Raisamo. 2022. Push-Poke: Collision based Direct Manipulation Technique for Plane Alignment in Virtual Reality. In *Proceedings of Graphics Interface 2022*. Canadian Human-Computer Communications Society, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 11 pages. https://api.semanticscholar.org/CorpusID:247517726

[39] Min Lin, Rich Goldman, Kathleen J. Price, Andrew Sears, and Julie Jacko. 2007. How do people tap when walking? An empirical investigation of nomadic data entry. *International Journal of Human-Computer Studies* 65, 9 (2007), 759–769. https://doi.org/10.1016/j.ijhcs.2007.04.001

[40] Yiqin Lu, Chun Yu, and Yuanchun Shi. 2020. Investigating Bubble Mechanism for Ray-Casting to Improve 3D Target Acquisition in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Washington, DC, USA, 35–43. https://doi.org/10.1109/VR46266.2020.00021

[41] 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

[42] I. Scott MacKenzie and Poika Isokoski. 2008. Fitts' Throughput and the Speed-Accuracy Tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. Association for Computing Machinery, New York, NY, USA, 1633–1636. https://doi.org/10.1145/1357054.1357308

[43] 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* 29, 4 (2017), 552–557.

[44] Halla B. Olafsdottir, Yves Guiard, Olivier Rioul, and Simon T. Perrault. 2012. A New Test of Throughput Invariance in Fitts' Law: Role of the Intercept and of Jensen's Inequality. In *Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers* (Birmingham, United Kingdom) *(BCS-HCI '12)*. BCS Learning & Development

Ltd., Swindon, GBR, 119–126.
[45] Xiangshi Ren and Xiaolei Zhou. 2011. An investigation of the usability of the stylus pen for various age groups on personal digital assistants. *Behaviour & Information Technology* 30, 6 (2011), 709–726. https://doi.org/10.1080/01449290903205437
[46] Bastian Schildbach and Enrico Rukzio. 2010. Investigating Selection and Reading Performance on a Mobile Phone While Walking. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (Lisbon, Portugal) *(MobileHCI '10)*. Association for Computing Machinery, New York, NY, USA, 93–102. https://doi.org/10.1145/1851600.1851619
[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 (dec 2004), 751–789. https://doi.org/10.1016/j.ijhcs.2004.09.001
[48] Da Tao, Jihong Zeng, Kaifeng Liu, and Xingda Qu. 2021. Effects of control-to-display gain and operation precision requirement on touchscreen operations in vibration environments. *Applied Ergonomics* 91 (2021), 103293. https://doi.org/10.1016/j.apergo.2020.103293
[49] Da Tao, Jihong Zeng, Dongming Shen, Xingda Qu, and Tingru Zhang. 2021. Touchscreen Operations in Vibration Conditions: Task Precision Requirement Matters. In *Proceedings of the 2020 2nd International Conference on E-Business and E-Commerce Engineering* (Bangkok, Thailand) *(EBEE '20)*. Association for Computing Machinery, New York, NY, USA, 60–64. https://doi.org/10.1145/3446922.3446933
[50] Nout C. M. van Zon, Clark Borst, Daan M. Pool, and Marinus M. van Paassen. 2020. Touchscreens for Aircraft Navigation Tasks: Comparing Accuracy and Throughput of Three Flight Deck Interfaces Using Fitts' Law. *Human Factors* 62, 6 (2020), 897–908. https://doi.org/10.1177/0018720819862146
[51] Cyril Vienne, Laurent Sorin, Laurent Blondé, Quan Huynh-Thu, and Pascal Mamassian. 2014. Effect of the accommodation-vergence conflict on vergence eye movements. *Vision Research* 100 (2014), 124–133. https://doi.org/10.1016/j.visres.2014.04.017
[52] 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)*. Association for Computing Machinery, New York, NY, USA, 657–666. https://doi.org/10.1145/1240624.1240727
[53] Alan Travis Welford. 1968. *Fundamentals of skill.* Methuen Publishing, London, UK.
[54] 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)*. Association for Computing Machinery, New York, NY, USA, 1639–1648. https://doi.org/10.1145/1978942.1979181
[55] Julia Woodward, Jahelle Cato, Jesse Smith, Isaac Wang, Brett Benda, Lisa Anthony, and Jaime Ruiz. 2020. Examining Fitts' and FFitts' Law Models for Children's Pointing Tasks on Touchscreens. In *Proceedings of the International Conference on Advanced Visual Interfaces* (Salerno, Italy) *(AVI '20)*. Association for Computing Machinery, New York, NY, USA, Article 56, 5 pages. https://doi.org/10.1145/3399715.3399844
[56] Charles E. Wright and Francis Lee. 2013. Issues Related to HCI Application of Fitts's Law. *Human-Computer Interaction* 28, 6 (2013), 548–578. https://doi.org/10.1080/07370024.2013.803873
[57] Savita Yadav, Pinaki Chakraborty, Lokesh Meena, Deepanshu Yadav, and Prabhat Mittal. 2021. Children's interaction with touchscreen devices: Performance and validity of Fitts' law. *Human Behavior and Emerging Technologies* 3, 5 (2021), 1132–1140. https://doi.org/10.1002/hbe2.305
[58] Shota Yamanaka. 2021. Utility of Crowdsourced User Experiments for Measuring the Central Tendency of User Performance to Evaluate Error-Rate Models on GUIs. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 9, 1 (Oct. 2021), 155–165. https://doi.org/10.1609/hcomp.v9i1.18948
[59] Shota Yamanaka. 2024. Relative Merits of Nominal and Effective Indexes of Difficulty of Fitts' Law: Effects of Sample Size and the Number of Repetitions on Model Fit. *International Journal of Human–Computer Interaction* 1, 1 (2024), 18 pages. https://doi.org/10.1080/10447318.2024.2303201
[60] Shota Yamanaka, Taiki Kinoshita, Yosuke Oba, Ryuto Tomihari, and Homei Miyashita. 2023. Varying Subjective Speed-Accuracy Biases to Evaluate the Generalizability of Experimental Conclusions on Pointing-Facilitation Techniques. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 317, 13 pages. https://doi.org/10.1145/3544548.3580740
[61] Shota Yamanaka and Hiroki Usuba. 2022. Computing Touch-Point Ambiguity on Mobile Touchscreens for Modeling Target Selection Times. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 186 (dec 2022), 21 pages. https://doi.org/10.1145/3494976
[62] Shota Yamanaka, Hiroki Usuba, and Homei Miyashita. 2022. Bivariate Effective Width Method to Improve the Normalization Capability for Subjective Speed-Accuracy Biases in Rectangular-Target Pointing. In *Proceedings of*

the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 211, 13 pages. https://doi.org/10.1145/3491102.3517466

[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 February 2024; revised May 2024; accepted June 2024